



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 172388

TO: Tekchand Saidha
Location: rem/2A65/2C70
Art Unit: 1652
Monday, December 12, 2005

Case Serial Number: 10/621826

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Saidha,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

THIS PAGE BLANK (USPTO)

From: Saidha, Tekchand
Sent: Wednesday, November 23, 2005 10:11 AM
To: STIC-Biotech/ChemLib
Subject: 10621826 - sequence search request

RECEIVED
NOV 23 2005
STIC/BIOTECH/CHM LIB

10621826

Please search all commercial and US patent data bases (including interference files) for:

1. SEQ ID NO: 2 and
2. SEQ ID NO: 3

Thank you !

Tekchand Saidha
Primary Examiner
Art Unit 1652, Remsen Bldg. E02A65 (Office),
Remsen Bldg. 02C70 (Mail Box)
(571) 272 0940

November 23, 2006

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: 12/1/05
Date completed: 12/12/05
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# 1 AA# 1
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: DBH
SEQUENCE SYSTEM: DBH
WWW/Internet: _____
Other (Specify): DBH

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
Remsen Bldg. 01 D86
571-272-2507

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC-Biotech-Chem Library Remsen Bldg.



THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 11:46:00 ; Search time 9332 Seconds
(without alignments)
11268.794 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgaggcgccatccacgag.....tcagctcgatccgacgagg 1850

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_sts.*

11: gb_sv.*

12: gb_un.*

13: gb_vi.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	256.4	13.9	22511	1	AF254925	AF254925 Streptomyces
C 2	242.4	13.1	85163	1	AY048670	AY048670 Streptomyces
C 3	242.4	13.1	292100	1	SC0939121	AL939121 Streptomyces
C 4	241	13.0	110000	1	AP006840_33	Continuation (34 o
C 5	241	13.0	110000	1	AP006840_34	Continuation (35 o
C 6	236.4	12.8	126933	1	DQ118863	DQ118863 Streptomyces
C 7	232.8	12.6	110000	1	BA000030_41	Continuation (42 o
C 8	231.6	12.5	110000	1	CP000010_06	Continuation (7 of
C 9	230	12.4	110000	1	BX571965_28	Continuation (29 o
C 10	227.2	12.3	110000	1	CP000031_23	Continuation (24 o
C 11	226.2	12.2	110000	1	AP006618_13	Continuation (14 o
C 12	223.8	12.1	2946	1	STMHUTH	M77841 Streptomyces
C 13	217	11.7	110000	1	BA000040_68	Continuation (69 o
C 14	216.8	11.7	49736	1	AF319998	AF319998 Stigmatel
C 15	215.4	11.6	1527	6	AR626602	AR626602 Sequence
C 16	215.4	11.6	6439	6	AR619418	AR619418 Sequence
C 17	214.6	11.6	110000	1	CR555306_34	Continuation (35 o
C 18	206.6	11.2	110000	1	BA000045_32	Continuation (33 o

C 19	198.6	10.7	203050	1	AL646071	AL646071 Ralstonia
C 20	191.8	10.4	301617	1	AE016911	AE016911 Chromobac
C 21	189	10.2	203050	1	AL646078	AL646078 Ralstonia
C 22	177.2	9.6	213732	1	AE001862	AE001862 Deinococc
C 23	176.8	9.6	10225	1	AE005048	AE005048 Halobacte
C 24	175	9.5	110000	1	CP000090_29	Continuation (30 o
C 25	173.4	9.4	10431	1	AE005773	AE005773 Caulobact
C 26	173.4	9.4	11099	1	AE004922	AE004922 Pseudomon
C 27	172.8	9.3	110000	1	AE014295_18	Continuation (19 o
C 28	172.8	9.3	349980	6	AX492785	AX492785 Sequence
C 29	172.8	9.3	349980	6	AX492786	AX492786 Sequence
C 30	172.8	9.3	349980	6	AX553952	AX553952 Sequence
C 31	172.8	9.3	349980	6	AX553953	AX553953 Sequence
C 32	171.4	9.3	11038	1	AE011796	AE011796 Xanthomon
C 33	171	9.2	110000	1	AE014292_09	Continuation (10 o
C 34	171	9.2	110000	1	AE017224_02	Continuation (3 of
C 35	170.8	9.2	10029	1	AE009324	AE009324 Agrobacte
C 36	170.8	9.2	13627	1	AE008290	AE008290 Agrobacte
C 37	170.8	9.2	110000	1	BA000012_68	Continuation (69 o
C 38	169.2	9.1	110000	1	CR555306_33	Continuation (34 o
C 39	164.2	8.9	110000	1	CP000009_12	Continuation (13 o
C 40	163.6	8.8	30390	6	CQ363732	CQ363732 Sequence
C 41	163.6	8.8	110000	1	AE017283_23	Continuation (24 o
C 42	163.2	8.8	1527	6	AR389675	AR389675 Sequence
C 43	161.8	8.7	1542	6	AX350741	AX350741 Sequence
C 44	161.6	8.7	9926	1	AE012259	AE012259 Xanthomon
C 45	161.6	8.7	110000	1	CP000050_31	Continuation (32 o

ALIGNMENTS

RESULT 1	AF254925/c	22511 bp	DNA	linear	BCT 01-JUL-2002
LOCUS	AF254925	Streptomyces maritimus	enterocin biosynthetic gene cluster,		
DEFINITION	AF254925	complete sequence.			
ACCESSION	AF254925				
VERSION	AF254925.1	GI:8926181			
KEYWORDS	Streptomyces maritimus				
SOURCE	Streptomyces maritimus				
ORGANISM	Streptomyces maritimus				
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomycetes.				
TITLE	Piel, J., Hoang, K. and Moore, B.S.				
JOURNAL	Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis				
REFERENCE	J. Am. Chem. Soc. 122, 5415-5416 (2000)				
AUTHORS	2 (bases 1 to 22511)				
TITLE	Piel, J., Hertweck, C., Shipley, P.R., Hunt, D.M., Newman, M.S. and				
REFERENCE	Moore, B.S.				
TITLE	Cloning, sequencing and analysis of the enterocin biosynthesis gene				
JOURNAL	cluster from the marine isolate 'Streptomyces maritimus': evidence				
PUBMED	for the derailment of an aromatic polyketide synthase				
REFERENCE	Chem. Biol. 7 (12), 943-955 (2000)				
AUTHORS	3 (bases 1 to 22511)				
TITLE	Hertweck, C., Jarvis, A.P., Xiang, L., Moore, B.S. and Oldham, N.J.				
JOURNAL	A mechanism of benzoic acid biosynthesis in plants and bacteria				
PUBMED	that mirrors fatty acid beta-oxidation				
REFERENCE	Chembiochem 2 (10), 784-786 (2001)				
AUTHORS	11948863				
TITLE	4 (bases 7073 to 7798; 12734 to 14128)				
JOURNAL	Xiang, L., Kalaizis, J.A., Nilsen, G., Chen, L. and Moore, B.S.				
PUBMED	Mutational analysis of the enterocin favorskii biosynthetic				
REFERENCE	rearrangement				
AUTHORS	Org. Lett. 4 (6), 957-960 (2002)				
TITLE	5 (bases 1 to 22511)				
JOURNAL	Moore, B.S., Piel, J. and Shock, D.				
PUBMED	Direct Submission				
REFERENCE	Submitted (12-APR-2000) Department of Pharmacology and Toxicology,				
AUTHORS					
TITLE					
JOURNAL					

FEATURES	College of Pharmacy, University of Arizona, Box 210207, Tucson, AZ	
	85721-0207, USA	
source	Location/Qualifiers	
	1..22511	
gene	/organism="Streptomyces maritimus"	
	/mol_type="genomic DNA"	
CDS	/db_xref="taxon:115828"	
	513..1004	
gene	/gene="encE"	
	513..1004	
CDS	/gene="encE"	
	/codon_start=1	
gene	/transl_table=11	
	/product="putative regulatory protein EncE"	
CDS	/protein_id="AAF81720.1"	
	/db_xref="GI:8926182"	
gene	/translation="MAQERRMQIGEVAERTGLSRTIRHYEEVGLIVPSARSKGGFPL	
	YTQADVSLMIVIRMKPUDFLEEMRDLLEITDRLGAQDDPPAPVERERLRERLDAYR	
CDS	KVADARCBETLRARLAEADFAATWRDGMNADARASASERATSGGFLRLWLTSADPG	
	LAG"	
gene	1749..2600	
	/gene="encF"	
CDS	1749..2600	
	/gene="encF"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="putative transcriptional regulator EncF"	
	/protein_id="AAF81721.1"	
gene	/db_xref="GI:8926183"	
	/translation="MDVEVLGHVTLKVGDERRTSPAKARQVLALLLVNANRMVSTDV	
CDS	LIEELWARPQRSALSALQTVFOLRLLKIRPDECPKOLLVTPNGYEPVPPENV	
	DLQYERHRLAQQAQSGENMAAVEQFRAALHIFAAPPLANVRGGMHLEAEIYRLSE	
gene	SRIGQVETRIAELEGRHVEVLSELGLASQCQHEGFHRQLMALRYCGRRRQALK	
	VFNRQSLVELGUEPSRELQMLQALLSADPRDLTAERTFFLRTTADPAPLQVPWP	
CDS	AASPWKV"	
	complement (2672..3163)	
gene	/gene="encG"	
	complement (2672..3163)	
CDS	/gene="encG"	
	/codon_start=1	
gene	/transl_table=11	
	/product="EncG"	
CDS	/protein_id="AAF81722.1"	
	/db_xref="GI:8926184"	
gene	/translation="MTTAGADPTQAEBSALTOAPHKPEETAALFAEIQOQFYGRHWRAM	
	DEGRVEDWTGPFMDPAFATNARPEPQGRGAEIARNAEAARQLOEKGILRRHCVTTL	
CDS	ELQQAQGLTLAKTYALITKTVVGGSRSELEFVCTCEDVLVRNENRWFIRHQVPRDDL	
	PKR"	
gene	3330..4937	
	/gene="encH"	
CDS	3330..4937	
	/gene="encH"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="putative acyl-CoA ligase EncH"	
	/protein_id="AAF81723.1"	
gene	/db_xref="GI:8926185"	
	/translation="MEPAPQAPTQPVIDARPEAGPRWETIGELIKDAALLHGDKFEL	
CDS	RCGKNTLSFSTDTTDRDLAQALIAQGVKGDVAVVMDNDVPLSWFAAIAKAGAIT	
	VPVNTRFGATLAHLKDSQAVRVLASPGCVPLARDVSGSIGCHPFCVARTLRELEADFQ	
gene	DRPVDGCLSAHADDTVNFOYTSGTTPGPKACMLSHDYWLRTANWIAVHSGLRPDDVV	
	LTAQAFSYMDPQWAKVCLMGVPLVLVPRESASGFWSVPOHRTATLYYLGSWPMLL	
CDS	YKQPHSGDRDHMRVLVCSGIPRDLHAHAFEDRWGAWREYVYSGTESGLDLIMPGEZ	
	ATVGSQWGYPPSPGREYIVADERQRPVQPGQIGELVGRGPMKGMWNNPDSTRAFR	
gene	GWWYHTDGLGRAEAGSVYHAGRLKMDIRGGENIAAEVESVLEAHPAVLAAALAGI	
	PDELFGELPKAFLOLRPGYRPTTATARSVLATRRHLAKFKVPAYVBFVDSFMTPSA	
CDS	RIQKRQLLRPGDDQRTGTGAFAADAWA"	
	4934..5710	
gene	/gene="encI"	
	4934..5710	
CDS	/gene="encI"	
	/codon_start=1	
gene	/transl_table=11	
	/product="putative enoyl-CoA hydratase EncI"	
CDS	/protein_id="AAF81724.1"	
	/db_xref="GI:8926186"	
gene	/translation="MSDVQQLSSADLTVTVDSVAVIAMSRAKLNALTRGMRDLA	
	ELLRHVGDSQARGVITGQAKAFSAQMSAGNDRVAGPEVDLLAEMLFNDITRAALITE	
CDS	VPVVAALGIAGVAGEMALSFDAIATAQAASISWPEVTGVLTVSNAAASLPLPRLAGA	
	PRALHLTMNARSALQALGLDDIVPEPDLIPAAIRLIHQWTPPGSSSTAHLRLM	
gene	RPSLRVAAAMAREGAARVQVQESGIIAQAGIATVLAGHEK"	
	complement (5856..7058)	
CDS	/gene="encJ"	
	complement (5856..7058)	
gene	/gene="encJ"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="putative beta-oxoacyl-CoA thiolase EncJ"	
gene	/protein_id="AAF81725.1"	
	/db_xref="GI:8926187"	
CDS	/translation="MTSGATYRDAAIPLTALWTSPPFARMGALSHLSLSDLAVAATGR	
	ALSERGIDRTIEGMVLGWTVPQPAIFYGATTLAARLGMFQVSGPMLNQACATVAAL	
gene	HQAAAVGTGVPOLIVTTRTNGPDLSPAGSQGPVRENWVRDSQDDPVTGQ	
	SMLATAEVAACGFTEDDLDAVLSECOYTAALSDDRTFORDFPMVGIEIEQDGLGT	
CDS	VRLLSDEGVRPVTADSAGQLPTVRPGSVHTAATQTHEPADGTGALVITTYTCARSPADR	
	GPVIRLVASGFARVEPARMPRALVPAAQAALHAAGLKIDTMDALSTNHPFAVNDLIYA	
gene	RETGVTLAEMNASGCSLLFGHPQPGTGLRTVTTELAHVLKARGGGGLGFTGCAAGDTAA	
	ALIIIVTD"	
CDS	complement (7073..7798)	
	/gene="encK"	
gene	complement (7073..7798)	
	/gene="encK"	
CDS	/codon_start=1	
	/transl_table=11	
gene	/product="putative O-methyl transferase EncK"	
	/protein_id="AAF81726.1"	
CDS	/db_xref="GI:8926188"	
	/translation="MTIDASPESAFEGEQSYEHRRLRYDAVVVHASNRLWRCR	
gene	KSRLLHEYNDWITPSRLEVGSGSYLDHCRFPAGTPELTLLDNLPLKFAAHLRR	
	YTPRCVQADILEPLFPVSGSFTSIANYVMHCLPEPPGGKQIVFKHLRALAPDGLV	
CDS	FGSTVLSQGVKHTLASRRFNILYQRQGSFHNQDHSVHGLHQALEEHFASHWIEVRGSV	
	ALFAARASQTAPRHPGASLDQR"	
gene	complement (7826..8626)	
	/gene="encD"	
CDS	complement (7826..8626)	
	/gene="encD"	
gene	/codon_start=1	
	/transl_table=11	
CDS	/product="putative ketoreductase EncD"	
	/protein_id="AAF81727.1"	
gene	/db_xref="GI:8926189"	
	/translation="MPAHTESPSVALVTGTSIGLISIAQALLQDGLKVIITGRDPGK	
CDS	LETVRSCLKMEDDQFDVAGYTCVDRDRDAVHDMVERCAKEHGEKVLVNNAGSGGGV	
	TAAIPDELWLDVINTLNSVFWTVREVRSSDAIKOPAGRIINTASYGKGQVPGYAP	
gene	YSAKAGVIGFTKALAKEFAAHGGPTVNAVCPGYVETPMAQDIRASAKLNNVTEQV	
	LADFEAKIPUGRYCTPEEVAGMVRKYLVPAAASVTAQALNVCGGIGSI"	
CDS	8815..10086	
	/gene="encA"	
gene	8815..10086	
	/gene="encA"	
CDS	/codon_start=1	
	/transl_table=11	
gene	/product="putative keto synthase alpha EncA"	
	/protein_id="AAF81728.1"	
CDS	/db_xref="GI:8926190"	
	/translation="MSHVVTIGLSVLARGSTDETEGFWKMITAGRSAIRRTAFDPTP	
gene	FRSQAGEVDLDPLACGFSRREVRRLDRASLLAVACARRAVERAGTISGTSIDPGRIG	
	VSVGNVAGSATSINENYVLSDEGQWLVDQKQSPHLFDYFVPGSLARETAAWAGAE	
CDS	GPVSVISAGPTSGIDALGACQLIQEGSADVMVAGATDAPVTPIAVACFDAIKATRR	
	NDDEPHAPRDFDRTRDGFALAEGAALVLEDSEHARRRGDIYGTITCYAAHSNAYHM	
gene	TGLSRDGNEMARAITTALDEARPTDPTAVGYNIAHSGSTKQNDLHETAEFAKALGEHAY	
	NVPVSSIKSMIGHSLGAIGAEVATCALIENGWVPTANLHEPDPDCLDLYVPLEAR	
CDS	DVLDDVVLTVGSGFGGFSAMIVEKWRGGHV"	
	10083..11306	
gene	/gene="encB"	
	10083..11306	

```
/gene="encB"
/codon_start=1
/transl_table=11
/product="putative keto synthase beta EncB"
/protein_id="AAF81729.1"
/db_xref="GI:8926191"
/tranlation="MSTDIVITGCVTAPTGLSVKEYNATLACTOGTAPLOEFDPTP
YTLRAGQITFEAKRHLPRLLPQTRVTRHALVAADRALEDQVDVGLGTVECGV
VTSNATGGFEFSHREMRQWLTQGRVSVYQCFWFAVNTGQJLSIRHMRGPGVVV
AEQAGGLDAIGHARTITRGRSLMTVGVSSFPDPMGWSHLSVSTATDPATAY
LPFQRAAGYVPGGGLLVNEDQASARGRTDRPYKITYAAATFDPAPCAQPTPSGL
RFDVLADALAGLPDAIDLVAADAGVADLDAEADALRAVFGYGVPCAPKTMVG
RLMAGGSPLDVVCALLALRDNVIPAVNVTVDDYQIDLVRDEPRTPCLSRALVLR
GAGFNSAVIVEHP"
11354..11602

Query Match      13.9%; Score 256.4; DB 1; Length 22511;
Best Local Similarity 51.4%; Pred. No. 4,1e-24;
Matches 726; Conservative 0; Mismatches 666; Indels 20; Gaps 5;

Qy 183 GCCGTGAGCTGGATCGCACATCATGATCTGGACACAGGCCCATGCGGTGGAGCGCGGC 242
Db 17831 GTCATAGAGCTCGACATGAACGTCAACGCTCGACCAACTTGAGGACCGCGCGCAGCGC 17772

Qy 243 GCGCGATTGCTTGCCTTCCCTCCGCGCGGACCGTGCCTGCGTCCGACCGCGGCTC 302
Db 17771 ACGCCCGTGGAGCTGTCGCGACCCGCTCCGCTCCGCGCTCCGCGGACGTGTG 17712

Qy 303 GCGCTGTGCATCCGCGGCGCGCCATGTCTACGGACTGACAAACCGGCTTCGGTCCCTT 362
Db 17711 GTGAAGTTCGTGACGACGACGTGTATCTACGGGTCAACACGAGCATGGGGGCTTC 17652

Qy 363 GCGAACCGCTGTATCTCAGGTGAGAAATGTCCGAACGCTTGAGGCCCAATCTTGTTCATCAT 422
Db 17651 GTCGACCACTCGTCCCGGTGTCCAGCGCGCGAGCTCCAGAGAACTCATCAACGCG 17592

Qy 423 CTGGCCAGCGCGTGGACCGGTGCTTACTTGACGACGCGCGCGCCATGGTTCGGG 482
Db 17591 GTCGCCACCAACGTGGGGCGGTATCTGACGACACACGCGCGCGGACCATATGCTGTCC 17532

Qy 483 CGTCTGTGTGTCATCGCTCAGGAGCTCCGCTGCGACGAGGAGGACCATCGCTCGCCTG 542
Db 17531 CGCATGCTGTGCTGGCGCGGGAACTCCGGGATCACCCCGGCGAATCTGGACAGCTG 17472

Qy 543 ATCGACCTGTCAATTCGAGCTCGCTCCCGCGCTTCCAGCGCGGACCGTGGCGCG 602
Db 17471 GTGGCGGTACTCAACCGCGGATCGTGCCTGATCCCGGAGAGGGCTCTTTGGGCACC 17412

Qy 603 TCGGTGACCTGACACCGCTTGGCATATGCTGCTCGCTCCAGGCGCGGAGACTTC 662
Db 17411 AGCGGTGACCTCGGCGCGCTGGCGCGGATCGCCCTGGTGTGGCGGGGACGTGGAGGCC 17352

Qy 663 CTGGACCGGACGAGACCGCGCTTGACCGCGCAGAAAGGCTCCGCGCGGACCGGCTGCA 721
Db 17351 CGCTAC---AACGCTGAGATATGCTCCCGGGCGGAGAGCCCTGTCCGAGCGCGGCTGCA 17296

Qy 722 ACCGCTCGATCTCTCCCATCGCATGCACTGCGCTGTGTCAACGAGGACCTCGCCATGAC 781
Db 17295 GCCGATGAGCTGAGCTACAGGATGGCTGCGCCCTGTATCAACGCGACGTCAGGACATGGT 17236

Qy 782 CGGATCGCGTGGTGAATGTCACGCTGCGCCATCTCGGCACTGGCGGCGGTGGCGTT 841
Db 17235 CGGCTGGGACCATATGTTCTCCAGCGCGCGCGGCTCGTGGAGCCGCTACCTACGAGT 17176

Qy 842 GACGCGCTGTCTGGGAATGTCTAGAGGCGCGGACCGAGGATGGGCGCGGCACTGTC 901
Db 17175 GTCGCGTGTGCTGAGGGCTTGGAGGATGATGAGAAACGTTTCGACCTTCGCGTGA 17116

Qy 902 CGACTCGCGCGCATCTCCGGAAGAAGACCGCGACGAGGCTGCGCGCGCGCGTGA 961
Db 17115 CGCGGTCAAGCGCACCGCGGACGCTCAGGTGGCTTGGCGGCTTGTGGAGGGGCTTGC 17056

Qy 962 CGGACGCGCGGGTGTCTCGGCACGTATTCGCCAGCGGAGGCTCGAC-----GC 1012
```

```
Db 17055 CGACTCGCACTGGCGGTCAACGAACTGACACCGAGCAGACCTTGGCGGAGAGATGGG 16996
Qy 1013 CGGCGATATCGGGACGGAGCGGGGAGGATCCCTACACCTTGGCTGCGCTGCCCTCC 1072
Db 16995 CACGTCGCAAGGCGGCTTGGTGGCGATCGAGGACGCTTACTCATCCGCTGACGCGC 16936
Qy 1073 CGAGTTCTCGGGGGGGCTTGCACACGCTCGCATGGCATGACCGGTGCTGACAGTGA 1132
Db 16935 GCAGATCTCTCGGTCCCGTGGTTCGATGTGTGACCGGATCGGGGCGACCTTCAGGACGA 16876
Qy 1133 GCTGAACCGCGTGCACCGCAATCCGGTGTTCGCGCCGATGGCAGCGTCCCGCCCTGCA 1192
Db 16875 GCTGAACCTCTCAACGACCAACCGATGCTCTGCGGAGG---AGGCGGAGGTGTTC 16819
Qy 1193 CGGGGGCAATTTTCATGGGCGACGATGTGGCGCTGACGCTCGATGCTGCGCACGCGCT 1252
Db 16818 CAACGGGCACTTCACGCGCAGTACGTGGCCATGGCCATGGACCACTTGAACATGGCCCT 16759
Qy 1253 CACGTTCTCGGGGGCTTGGGAGCGCCAGATTGCAGCTCTGACAGATGAAGGCTGAA 1312
Db 16758 GGCACCGTGAACCAATCTCGCAACCCGCGGTGGACCCCTTCCTGGACAAAGCAACAG 16699
Qy 1313 CGGTGGGCTGCCCCCTTCTCCACCGGGGCCCCCGGGTTGAATTCGGGCTTCATGGG 1372
Db 16698 CAACGGGCTGCGCGCTTCTGTGGCGGAGATCCGCGACTGCGCTGGGCTGATGGG 16639
Qy 1373 CGCAAGGTGACGCGACCGCGCTCTGCGCGAGATGCGAGCC---ACGGGACCTGCTC 1429
Db 16638 CGGCAGTTATGATCGCGCTGATCACCGCGGAGACCGGACCTTCCGATCCGATGTC 16579
Qy 1430 GATCATTCGATCTCCACGACCGCCCAATCAGATGTGCTGCTGGGACCATCGC 1489
Db 16578 GGTGAGTCCCTCAAGAGTACGCGGAGCTTCAGGACATCGTGTCTTCGAGATTCGTCG 16519
Qy 1490 CGCGCGCTCTGCGCGGAGAGATCGACCTTGGCGGAGATCTTGGGATCTCGCTCT 1549
Db 16518 CGCGCGCGCGCGCGGAGGTACTCACCACGCTGCTACGTGGTGGCTTCGAGTGTCT 16459
Qy 1550 CTGTTTGCACAAAGTGGAGCTGCGCTGCG 1581
Db 16458 GTGGCGCTGCCAGCGCTCGACATCCGCGGCG 16427

RESULT 2
AY048670/c 85163 bp DNA linear BCT 05-NOV-2002
LOCUS Strepomyces globisporus enediylne antitumor antibiotic C-1027
DEFINITION biosynthetic gene cluster, complete sequence.
ACCESSION AY048670
VERSION AY048670.1 GI:24575076
KEYWORDS
SOURCE
ORGANISM
Strepomyces globisporus
Strepomyces globisporus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomycetes.
REFERENCE
1 (bases 1 to 85163)
AUTHORS Liu, W., Christenson, S.D., Standage, S. and Shen, B.
TITLE Biosynthesis of the enediylne antitumor antibiotic C-1027
JOURNAL Science 297 (5584), 1170-1173 (2002)
PUBMED 12183628
REFERENCE
2 (bases 1 to 85163)
AUTHORS Shen, B., Christenson, S.D., Liu, W. and Standage, S.
TITLE Direct Submision
JOURNAL Submitted (27-JUL-2001) Chemistry, University of California, Davis,
One Shields Ave, Davis, CA 95616, USA
FEATURES
source
1..85163
/organism="Strepomyces globisporus"
/mol_type="genomic DNA"
/strain="C-1027"
/db_xref="taxon:1908"
complement(8..658)
CDS
```

/note="ORF (-7) "
/codon_start=1
/transl_table=11
/product="hydroxylase-like protein"
/protein_id="AAL06648.1"
/db_xref="GI:24575077"
/translation="MGMTGARPRVLVVGAGLAGTATAIRLHFARRPLEVVLERRAA
YRAGVAYHRDQWDFVNIQAGMSVFREDVLDIFNMANOADRDWPRRWSKWF
TEQGPAPRRIFOYLDARLVEARESCPGVVLVEADGEALDASHDPCFSTVVRGLTP
YLTEGLRPGPLDADHVLATGLLEKEPPFAAGVAGHSFVRNPISAPGIL"
complement(930..1478)
/note="ORF (-6) "
/codon_start=1
/transl_table=11
/product="viral infectivity potentiator-like protein"
/protein_id="AAL06649.1"
/db_xref="GI:24575078"
/translation="MPRIPTHITDTAPAGGDLRLRLEKFRGLRVNLIHGMASPVVL
ETVAITGVAEHEGTFOARTREAIALAVGADVACAYCOAAHTVS AKVAGFTLEETVAI
RRGTPGDVLEALVQVAREIAGEVGEASDASWNAVAQGWTDTELAEVFVHAVNLY
TNVFNHARTEIDVPGVPDIDS"
complement(1649..2713)
/note="Orf (-5) "
/codon_start=1
/transl_table=11
/product="truncated methionine synthase-like protein"
/protein_id="AAL06650.1"
/db_xref="GI:24575079"
/translation="MTIATEPIGSLPRSATLLHALAAHQGTLDATDLAKQOQAVAD
TLRLERGVFLVDGSKPSFATYVTLGDLTSPDGAVIPFADGHTROLPSITQGP
FRYQRAETYLRAARQLTDRPLKQAVIAPSAISLIPATPIEGYPRQFRLDADAE
ADTRGCDAGAHVQVLDFTGRLLKQVLDLDDPFIALNNEVLGRFSAETARIGV
HCPGQDQSTHSLIDYELLPKLFOLKAGNVLELAGADPERVLSIVRDHLPPAA
RVLGVDTPDIPVEPQVDRILLARVIPVEQLGTCDCCGFSPPADTSTTRDLA
FAKIEARQGTALAEVILG"
complement(2851..3237)
/note="ORF (-4) ; similar to SCP1"
/codon_start=1
/transl_table=11
/product="transcription factor"
/protein_id="AAL06651.1"
/db_xref="GI:24575080"
/translation="MSSLTDSGLVHEAKVWVLEDPGLDYVQALDKTPRRKNKPRY
ARDGRMIGYTELGADAEADPDSGLYRERVFFLLPHDRDSDPEGVYRQAGAEAVDPRT
IENRVGEKTPRQLGTSSTVAATGS"
complement(3442..4971)
/note="ORF (-3) ; similar to SCP1"
/codon_start=1
/transl_table=11
/product="putative primase"
/protein_id="AAL06652.1"
/db_xref="GI:24575081"
/translation="MSSASGKPRFDTAAQAQMLAETTPAPAPLIPVQAAPAAVAA
TAWAGIIPASLTDGNAGLFLVRLYRDQFRHVEGLGWYSWDGYRWKRAGGEKAAALWAA
GEMAEPPGSDPRLGFLTDRELLHHKRTLTSTGMKALLTQAKASPDLSLDPDLTGDGP
YALCTPGVDLNRKMRKDPTRDHSRATSASPODIPTPRWRHFLEDTFGSDAERG
EMIDFLHLLAGYSTGDCQAVLPFLHGCKNGKSVLLDVMIOILGDYADAAAPGFLM
DRGAYSHSLTELHGRRLIVCSLKNDRPDEARVRLITGGDKIKARRMRODYRSF
TPYHKLWLLGNPEVSTGGFWRRLIRLLPFERIVDERKIDNLKAVELVQDSGPGIL
HWLTGARYLATRDTPGDVRIRIATSAVANTEDHIGRFLEACCLHDPENSLRVEQ
GLLYTSTYWCASHSEGIPTGTAFAFATRVQRVGLASPADMIDKSNRKFYPNLALAA
E"
5982..7478
/note="ORF (-2) "
/codon_start=1
/transl_table=11
/product="glycerol phosphate ABC transporter"
/protein_id="AAL06653.1"
/db_xref="GI:24575082"
/translation="MRALPGSGLRMAGDAPTAAGNRCCPTPIQRPLILRQTLQAP
EANKRSPQQRERIVLDVWNLNPPYTDGDFLAPLRERAEEFRAHPRYRVGIN
GHDFWTIPEKVARATEGRPHIAGYATDSQLARDRDPDGKPVFTSVAAALGRTE
ILGHPVVVEDLDPVVRDSYSGGELVSLPLTVTMTLCYANSSLLARAGVPELPTWDE

CDS

CDS

CDS

CDS

CDS

VEAACAVASVDGPGHGHTIANDGWVFOQAVALONGVLTQDNRRSGSATTVDTVTS
EMLDVWRWTHLHERGHLYTGQPSDWGGAFFAFOOKVAFPTDSSKAARELIOAGAO
AGEFVAVPLPRNAKAPVQPSGSLWLAAGLDTQDGLLALTOYLISPAANAOW
HRTNGFVPTGAAGELLEATGWFDRRPOQRVAGEQLKASDRSPAALGALLGDFPAVNE
VITTAAMDDVLRSGADPAFAEAGAAQQLLDAYNARNRSGSGTSPSAV"
complement(7573..9927)
/note="ORF (-1) "
/codon_start=1
/transl_table=11
/product="uvrA-like drug resistance pump"
/protein_id="AAL06654.1"
/db_xref="GI:24575083"
/translation="MYGQSENSVSTPTAGAGTSGPGDGIRIAGARIHNLKDVSL
TLPRNLTFTVTVGSGKSSIVFTVAEQROLNSTFSWIRNOLPKYERPOAEATE
NIPTVIVQKPVGNARSTVGTWTVQPMIRALFAFEGTKDGPSSLGVSASFND
PQWCFCDCGLGSIADLDLMDLRKSLDDGAVLPEYKVGSPDWIWKASGRDPA
KPIAEYSABELDILLRGTKGTIKTKSTFQNYEGLADRFEKLNKRDLSALSDDRK
RIEVERFTDVCPSRGARLNAAALDVRIDGKNIADYSMEVRDLTEVLAVTEPAA
APLKAARTALERIVSIGLGLTDLDRPTADLSGEGQLKMRHLGSLAGLTIFYDE
PSLKGPRDVRNLRLRLDKGNTVLVVEHDPDVEIADHIVDVGPAGVHGGEIV
FEGSFALKRKARTGTEALRRGARVKVEEVRPPTGELTVENADLHNLKQVSVAVPTGVL
TATVGAAGSKSLISGAFMEHPDAVFDQSAIAASSRSTPVSYLGLMPLRKLFAK
ETGANASLFSNKGCEBCECCQGVIIITELAFMDPVTTHCGVCGRAPKEVLEBHLR
GSIADVLELPDAVEFFTEKALLPKLRALVDGLGLSLGQPLSLSGGELQRIKL
ADQLHRTGVYVLDDEPTTLHMGDVTLLKLDLGLVEAGNTVVVIEHNLVDVQADWI
IDLPGDGRGGEIVFTGTPKDLLAADSSLTGTLYRLRHLKAQAG"
complement(9982..11349)
/note="ORF0"
/codon_start=1
/transl_table=11
/product="Na/H efflux pump"
/protein_id="AAL06655.1"
/db_xref="GI:24575084"
/translation="MAVLAATSPGCEPLTVFLIQVLLLACAYGLGRGLGTRIGLPL
VGLTAGVLGLPTLQGISPLSGRUFFPEDIQAHLDDAFQFGVGLLVAIAGQFDP
RLRKRGLAARVSLAGLILPLIGIATGYLVPASLIADSGERGFALFLGVAMCVTA
LPVIKTLADNLTHRNQQLITAAAVFDADVGMLLIALVTALASGAAGGPPVLTWTA
WTVFVAACAVGSPIGRRLSRTGDSRPVSATVGVAVVLYGALTAAAGMEALFGA
FVAGATLLRHIAVRPLAPLTMAYAPVPLGSLGVLGRLTALAEPSVLLTGLGVLL
VATFGPAGAVAAASGMSRYEGLALGAGNRMIEVIALVGLRIGLVDITVTFII
ILVVALITTSVTAPILRWASSRIVLEEDETERDRLAGWMTEPALSGGPAKPSAREBK
TPDTS"
complement(11351..12835)
/note="ORF11"
/codon_start=1
/transl_table=11
/product="alkylhalidase"
/protein_id="AAL06656.1"
/db_xref="GI:24575085"
/translation="MDVSAQYDVIVVGGPGAGSTVSTLVRKGRHVLQLEKETPPRYO
IGESLAPSTVHGIAHLGWSDELKKAFTIKHGSTFKWGANPEPTWTFDAVSRMPGA
TGAYOVERMKPQIILLDNARRHGVRENSDVLVDLKAEDGRVGRVYRDSREHRE
VGRFVVDASGNTGGYKKGAKRESYPPFRNLALFGYFNGGRKLPKPNKSEIFCTIF
EHGWFIYIPLSELTSGVAVNRDSASLVQGDPEKAMESFIAACPLIAEKLSAIRT
EGYGBLRVKGMSYNTKFWAPMALVDGACFVDFVFSVHLATYSGLLSAARSLN
SLCDGSVDTSAPEEFEARVLEYGFYFLGVFDMHSDSYFWOARKISNTADS
LQSFVLIQGVSDDEFADTARQRSEASDEMAGMTHHTAAGDADGPHGAVPGKVGR
LMQSAALQARALLGDEAGNERPVROGGLAASDTGLGWNVPGAR5"
13012..14079
/note="ORF10"
/codon_start=1
/transl_table=11
/product="dNTP-glucose synthase"
/protein_id="AAL06657.1"
/db_xref="GI:24575086"
/translation="MKALVLSGGSGTTLRFISVAMPKQLVPIACKPVLEVLDNIRNL
DIKEVAIVGDWQAQEIIEAMGDSGRFLRLTYIRQOPLGIACHVKLARFLDDEDFV
LYLGDIMLDLSAQAGHFLHTRPAARIVVRQVDPRAFGVIELDGRVLRVLEKPR
EPDSILAAVGVYFETADVHEAVDAISPRSEGELEITDAIOMLLEQGLPVBAGRVTYDW
DKTVRDEVECHNRMLRLALQVSEVDEPSELVGVAVVEGARVSVSVGVGPAVIG
AGTVSDSQIGPYASIGRRTCTVRASRUSDIVLDDASILAVSLHGLSLRGARIAPG
ARGEARHLRVGDHVIETAA"
complement(14212..14643)

```

/note="Caga"
/codon_start=1
/transl_table=11
/product="C-1027 apoprotein"
/protein_id="AAL06658.1"
/db_xref="GI:24575087"
/translation="MSRHHMSRRASRGFWAVASIGLAAAAQSAFAAFAPSVSPASG
LSDQGSVSVSGAAGAEYTYIAQCAPVGQDQACNPATATSFTTDASGAASFVVRVY
SYTSGTPGTGVSGLVGCATACNLGAGNSGLDLGHVALTFG"
complement(14690..15922)
/note="ORF8"
/codon_start=1
/transl_table=11
/product="amino transferase"
/protein_id="AAL06659.1"
/db_xref="GI:24575088"
/translation="MSCGPRVRLDSLPTFSEPLHVCRPNVGSRDRLMERIDGALERL
WFTNDGFLVRFEARVAELQVRHCVAVSNATTGIQVAALKALGIGPGEDEVIPSF7FW
ATAHLWDIGAVPVFCBDEETGPADVAHVRLIGFTRAILDLVHVGGRPAVIDELTL
LAAEHLGLLDPAAAFHAFSGTAKSPKXIFGCGTAIFGQATKFNVSPEGGAIVTDDAL
ADRLRMHRCQLGNAAEHTTSGTGVARMEHISAAMGLTSLAESADHFTAINRRNVRYLSP
YLDLGPVGRVRRQDPSNELSNCOVVYIEVDVAVRAGLHRDELQAVLQRHNVLAARYSPG
CHCEPRSDLAHAPDPLPKVEALTERTVLSTPTGTAVGFEVGRVCRILRAVDGSA
VPEIHSTEGDAGGRPAR"
complement(15919..16653)
/note="ORF7"
/codon_start=1

```

Query Match	13.1%;	Score 242.4;	DB 1;	Length 85163;
Best Local Similarity	50.3%;	Pred. No. 1.8e-22;		
Matches 749;	Conservative	0;	Mismatches 706;	Indels 33; Gaps 5;
Qy	227	CGTGGCGAGCGCGCGCGGAGTCTCTCTGTCGCCCTCCGGCGCGACCGGTGCGCGTCG	286	
Db	40901	CGTGGCGGAGAACGCGGACGCTGACGTACCGGCCGAATCATGCGGAGGCCAGAA	40842	
Qy	287	GTCGGAAGCGCGCTCGCGCTGTCTATCGCGAGCGCGCCATGCTTACGGACTGCACAAC	346	
Db	40841	GAGCCCGGAGATCTTCCGAAGGATCCCGGAAACAGAAATCCCATCTACGGGTGACCAAC	40782	
Qy	347	CGGCTTCGCTCCCTTTGCGAAACCGCTCTGATCTCAGGTGAGATGATTCGCGAACGCTGCAGGC	406	
Db	40781	CGGGTACGGCGAGATGATCTCATGCAAGTCGCAAGTCGAAGGAAGTCGAACTGCAGAC	40722	
Qy	407	CAATCTTGTCCATCATCTGGCCACGCGCGCTGGGACCGGTGCTTGACTGGAACGACGCGCG	466	
Db	40721	CAATCTCGTCCGTAGCACAGCGCGGAGTCGCTCGCTGTTTCGCCGAGGACGAGGCGCG	40662	
Qy	467	CGCCATGGTCTCGCGGCTCTGGTGTGATCGCTCAGGAGAGCCTCGGTGCGAGGAGGG	526	
Db	40661	GCGGATGCTGCGCGCCCGCTGAACCCCTCGCAAGGGCCACTCGCGGTGCGGCCCAT	40602	
Qy	527	GACCATCGCTCGCGCTGATCGACTGTCAATTCGAGCTTCGCTCGGCGCTTCCCAGCGG	586	
Db	40601	CATCTTCEAACGCTCGCGCAGTACCTGAAACGAGGGCATCACCCGGGCCATACCGAGAT	40542	
Qy	587	CGGCAACGTTGGGCGCTCGGGTGAACTGTACAACGCTTGGCATAATGGTGTCTGTGCTCCA	646	
Db	40541	CGGGTCACTCGGGGCGAGGGGAGCCTGGCTCCCTCTCCCACTGCGGAGCACCCCTCAT	40482	
Qy	647	GGGCGGGGAGACTTCTCTGACCGGGAACGGGACCGGGCTTGACGGCGCAGAGGGCTCGG	706	
Db	40481	CGGAGAGGGCTACGTCTG---CGGACACGACCGCGGTGGAGACCGCCCGAGGTGCTGGC	40425	
Qy	707	GGCGGACGCGCTGCACACCGCTCGATCTCTCCATCGCGATGCACATGGCGGCTGGTCAACGG	766	
Db	40424	CGAGCGGGGCATCGAGCGCTCGAACTGCGCTTCAAGAGGGGCCCTCGCACTGATCAACGG	40365	
Qy	767	GACCTCGGCCATCACCGGATCGCGCTGGTGAATGCTCAGCCTCGCGCCATCTCGGCAA	826	
Db	40364	CAGTCCGGGATGACCGGCTTGCGGTCTCCCTGGTCTCGGACGGGCGCTTGGAGCAGGCCCA	40305	
Qy	827	CTGGGCGGTGGCGTTTGAACGGCCCTCGCTTCGCGGAATGTCTTGAGAGGCGCCGACCGAGGCATG	886	

Db	40304	GCAGGCCGAGATCGTGACGGCTCTGTCTCATCGAGGCGGTACGCGGATCGACGAGCCCTT	40245			
Qy	887	GGCCGCGGCACTGTCCGACT---GGGCGGATCCGCGACAGAGGACGCGCGACGAG	943			
Db	40244	CTTCGCGGAGGGGACGACATAGCCCGCCGCGACGAGGGCCAGATCGACACCGCGCCAA	40185			
Qy	944	GCTCGCGCCCGCGTGTGACGCGACGCGGGGTGTCCCGCACGTCAATTGCCGAGCGGAG	1003			
Db	40184	CATGCGGGCCCTGATGCGGGCAGCGGACTGACGCTCGAGCAGCGGACCTTGGCCGAGA	40125			
Qy	1004	GCTC-----GACGCCGGCGNATTCGGGACGAGCGGAGCGGGGGGAGGA	1048			
Db	40124	ACTCCAGAGGACAAGGAGCGCGGCAAGGACGTCCAGCGCTCGGAGATCTACTCTCAGAA	40065			
Qy	1049	TGCTTACAGCTCGGCTCGCTCCGCAAGTTCTCGGGCGGGCTTCGACACGCTCGCATG	1108			
Db	40064	GGCCTACTCGCTCGGGCCATCCCCCAGGTGTCGGGGCGGTGCGGACACCTTGTACCA	40005			
Qy	1109	GCATGACCGGGTCTGACATCGAGCTGAACCGGTGACCGACAATCCGGTGTTCCTCGCC	1168			
Db	40004	CGCGGGCACAAGCTGGCATCGAGCTCACTTCGGCCACGACACCGCTCTT-----	39951			
Qy	1169	CGATGGCAGCGTCCCGCCCTGCA CGGGGGCAATTCATGGGCCAGCATGTGGCGCTGAC	1228			
Db	39950	CTTCGAGGGCAAGGAGATCTTCCACGGGCGAACTCCACGGTCAGCGGATCGCGTTGC	39891			
Qy	1229	GTCCGATGCGCTCGCCACGGCGCTGACCGTTCTGCGGGCCCTTTCGGAGCGCCAGATTGC	1288			
Db	39890	GATGGACTTCGTGACCATCGCGCTCACCCAGCTCGCGGTCTTCGGCGGACGGCAGATCAA	39831			
Qy	1289	ACGTCTGACAGATGAAGGCTGAACGGTGGGCTGCCCGCCCTTCCTCCACCGGGGCCCGC	1348			
Db	39830	CGGGTCTTGAACGGGCACCTCAGCTACGGCTTCGGGAGTTCTCGTCTCGGGGACCC	39771			
Qy	1349	CGGTTGAAATTCGGGTTTATGGGGCGCA CAGGTGACGGCGACCGCGCTCTTCGGCGAGAT	1408			
Db	39770	GGGCGCTGCACAGCGGATTCGGCGGGCCGACGATACCGGCGCACCGCACTGGTGGCGGAGAA	39711			
Qy	1409	GCGAGCCACGGGACCTCGCTCGATTCATTCGATCTCCAGACGGCGGCCCATCAGGATGT	1468			
Db	39710	CCGGACGATCGGGCCGGCCAGCACCCAGAGCGTCCCGTCCAACGGCGCAACACGAGACGT	39651			
Qy	1469	GGTCTCGCTGGGACCATCGCGCGCGCTCTGCGCGGAGAGATCGACCGTTGGGCGGA	1528			
Db	39650	GGTGAGCATGGGCTGATCTCGGCCGCGCAACGCCCGCGGGTCTGTTCGACACACAA	39591			
Qy	1529	GATCCTTGGCATCCTCGCTCTCTGTCTTGCA CAACTGCGGAGCTGCGCTGCGGAGCGG	1588			
Db	39590	GATCCTCGCG-----GTGGAGTACTTGGCGCGCCCGACGGCGGTGCGACATCTCGGCGC	39537			
Qy	1589	CCTAGACGGGTGTCTCCCGGGGAGAGCTGTGTGACGGCCCTGCGGAGCAGTTCCTCC	1648			
Db	39536	GTTCGACGGGTTGAGCCCGCGCGGAGAGGCCCATGTCACGAGCGGTGCGCTGGTTCC	39477			
Qy	1649	GCGCTTTGAGACGACCGGCCCTCTGGGACGAGAAATTCGCGCGCTTC	1696			
Db	39476	GACGCTGGGCTCGACCGGTCATGCGCGACGACATCGAGCTGTGTCG	39429			
RESULT 3						
SCO939121						
LOCUS	292100 bp DNA linear BCT 16-APR-2005					
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 18/29.					
ACCESSION	AL939121 AL031317 AL160431 AL161691 AL161755 AL161803 AL353672					
VERSION	AL356832 AL389898 AL450450 AL451102 AL589164 AL645882					
KEYWORDS	AL939121.1 GI:24429533					
SOURCE	Streptomyces coelicolor A3(2)					
ORGANISM	Streptomyces coelicolor A3(2)					
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;						
Streptomycineae; Streptomycetaceae; Streptomyces.						
REFERENCE	1					

AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.B., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Lake, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

JOURNAL Nature 417 (6885), 141-147 (2002)

PUBMED 12000953

REFERENCE 2 (bases 1 to 292100)

AUTHORS Bentley, S.D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk

COMMENT On or before Oct 30, 2002 this sequence version replaced gi:3449234, gi:7288050, gi:7320987, gi:7321265, gi:7649562, gi:8218190, gi:9367445, gi:11544744, gi:20520684.

FEATURES

source 1..292100

/organism="Streptomyces coelicolor A3(2)"

/mol_type="genomic DNA"

/strain="A3(2)"

/db_xref="taxon:100226"

97..103

111..3596

/gene="SCO4654"

/note="synonyms: rpoB, SCD82.26"

111..3596

/gene="SCO4654"

/note="SCD82.26, rpoB, DNA-directed RNA polymerase beta chain, len: 1161 aa; highly similar to SW:RPOB MYCTU (EMBL:L27989) Mycobacterium tuberculosis DNA-directed RNA polymerase beta chain (EC 2.7.7.6) RpoB, 1178aa: fasta scores: opt: 5849 z-score: 6361.5 E(): 0; 75.7% identity in 1169 aa overlap. Contains Pfam match to entry PF00562 RNA_pol_B, RNA polymerase beta subunit and match ot Prosite entry PS01166 RNA polymerases beta chain signature"

/codon_start=1

/transl_table=11

/product="DNA-directed RNA polymerase beta chain"

/protein_id="CAB77428.1"

/db_xref="GI:7248348"

/db_xref="GOA:Q9J0L0"

/db_xref="InterPro:IPR007120"

/db_xref="InterPro:IPR007121"

/db_xref="InterPro:IPR007641"

/db_xref="InterPro:IPR007642"

/db_xref="InterPro:IPR007643"

/db_xref="InterPro:IPR007645"

/db_xref="InterPro:IPR010243"

/db_xref="UniProt/Swiss-Prot:Q9J0L0"

/translation="MAASRASTANTNNASTAPLRISPAKIEPLEVPNLLAQTES FDMGLNDANKARVESAGQDVPVTSGLLEEFIEESIPDFSGSNLFRDHRFEP RVNSIDCKORDFTYAPLFTVAFBNTNNEGEIKSQTVFMGDFPLMNTKGTFFVINGTE RVVSLQVLRPGVGVFDSIDKSDKI FSAKII PSRGAWLEMEIDKRDMDGVRIIDKR KQSVTVLLKALGWTTEOILREFEVESMRATLEKHDTQGGDDALDIYRKLRCPEPT REAAQTLENIYNPRYDLAKVGRYKVKKLGADLEPQDAGVLTDDVIATIKYVLK HAGETVSGSGREIVVETDDIDHFGNRRIRNVGELLQNVQRTGLAMERVRERMTT QDEBAPFTPTLINRPVVASIKKEFQSQLSQFMQNNPLSGLTKRRNALPGGLS RRVAFVRDVHSHGMCPIETPEGENGLIGLSASYRINPFGFIETPYRKYVVG QVTDVQVLADEEDRVIAQNALGDMRFAEVLVRRRGVEDVYVPGDDVMD VSPRQVSYATAMIPTLEHDANRALGMANMRQAVPLIKESPLVCTGMEYR5AADA GDVVKAKAGVQVQSADYTTTNDCTYITTLAKESRSNQTQSVNQKVIYAGDRI IEGCVLADGPATENGEMALGNLLVAFMFWEGNEDITILISQRLVQDDVLSIHIEE HEVDARDTKLGPETEDILPNVSEVLADLDIEDIRIGAVFVAGDILLVGKTPKGT ELTPTEERLRAIFEKAREVRDTSKYPHGIEIGKVIQVRFDRREGDELPPGNQLVR VTYAQRKRKTGDGKLGRHGNKGVI5KINPIEDMPFLEDGTPVDIILNPLAVPSRNP

GOVLEIHLGLWASRGMDVSGLADEWAQRLOVIGADKVEPGTNTVATPVPDGGAREDELALG LIQHTYIPNRDGERMVLPSGKARLEFDGRSGSPPEPISVGVMYILKLHLHVVDDKLHARS TQSGYMITQOPLGGKAPQGGQRGEMEWALEAYGAYALQELLTISKSDVGTGRVKVY EALVKGNEIPPEGPSPFKVLIKEMOSLCLNVEVLSDGMSIEMROTDIEDVFRAREEL GIDLSRRRESSVEEV"

411..3356

/gene="SCO4654"

/note="Pfam match to entry PF00562 RNA_pol_B, RNA polymerase beta subunit, score 1264.40, E-Value 0"

2712..2750

/gene="SCO4654"

/note="PS01166 RNA polymerases beta chain signature"

3676..3682

3694..7593

/gene="SCO4655"

/note="synonyms: rpoC, SCD40A.01, SCD82.27"

3694..7593

/gene="SCO4655"

/note="SCD82.27, rpoC, DNA-directed RNA polymerase beta' chain (fragment), len: >1058 aa; highly similar to SW:RPOC MYCTU (EMBL:L27989) Mycobacterium tuberculosis DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) RpoC, 1316aa: fasta scores: opt: 5139 z-score: 5536.5 E(): 0; 72.9% identity in 1067 aa overlap. Contains Pfam match to entry PF00623 RNA_pol_A, RNA polymerase alpha subunit. Contains also possible coiled-coil region approx. at residues 168..220

SCD40A.01, rpoC, DNA-directed RNA polymerase beta' chain (fragment), len: >279 aa; similar to C-terminal region of SW:RPOC_BACSU (EMBL:L43593) Bacillus subtilis DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) RpoC, 1199 aa; fasta scores: opt:801 z-score: 911.4 E(): 0; 53.0% identity in 247 aa overlap"

/codon_start=1

/transl_table=11

/product="DNA-directed RNA polymerase beta' chain (fragment)"

/protein_id="CAD55212.1"

/db_xref="GI:24429534"

/db_xref="GOA:Q8CJ71"

/db_xref="InterPro:IPR000722"

/db_xref="InterPro:IPR006592"

/db_xref="InterPro:IPR007066"

/db_xref="InterPro:IPR007080"

/db_xref="InterPro:IPR007081"

/db_xref="InterPro:IPR007083"

/db_xref="UniProt/Swiss-Prot:Q8CJ71"

/translation="MLDVNFDELRIGLATADDIRQMSHGVEKKPETINTYRTLKPEKD GLFCEKIFGTRDWCYCGKYKVRFGKIIICERCQVEVTRAKVRRRMGHIELAAPVT HIWYFKGVSRLGYLLDLPKLEKVIYFAAYMITFVDEERTRDLP5LEAHVSVERQ OIQBRDSDLEARAKLETDLAEAEAGKADVRKVRGEAREMKOLRDRAQREIDR LDEVNRFRNLKVDLEGBDELLYRELDRGTYPDGSMAAALQKRLSEFDLDEAER LREIITRGQKKTRALKRKVVSFAFQTSNPGMWLDCCVPVPPDLRFMWQDGGR FATSDLNDLYRVRINRRNLKRLDIGAPEIIVNNEKRMLOEADALFDNRRGRPVT GPONRPLKSLDMLKQKGRFRQNLGKRVDSARSVIVVGPOLKLCGLPKAMALE LKPFVFKRLVDLNLHAQNIKSAKMWVERGTVTVVLEEVIAEHPVLNRAPTLRGLG IOAFEPOLVEGKAIQIHPLVCTAFNADFDGQMAVHLPLSAEAAQAEIIMLSNNIL KPADGRVTNPTQDMLGLFLETTDEGRSPKGEGRFSGSSAEAINAFADGLTLQAK IDIRFPVGTIPPRGFPPPAAREGEPEWQGDFTLTKTLGRALNFLEALDIPVDDKXI VKGQSEIIVNDLAERYPKVIVATLNDLKAAGFFWATRSQVTVASDIIVVDDAKKI VKGEGEDQKQYVERGLITKEERTQELIAITWKATNEVAEMNDNFPKTPVSMVY NSGARGMMOMROIAGMRGLVSNAKNETIIPRIKASPREGLSVLEIVSTHGARKGLA DTLATADSGYLTRRLVDSQDVIIBREDCGTERGLKPIATRADGTLKARQVETS VYARMLAEDVVIDGKVIAPANVLGDVLIDALVANGVEEVKTRISILTCESQVGTAMC VYKSLATGLKVIDGEAVGIIAAQISIGEPGLQTLNRTPTGTVAGDDITQGLPRVVELF EARTPKGPAVISEASGRVRIETEETKTKI VTPDGDSDETAFFI5SKRALLVGGDHF EVQKLTGATNPDLVRLIIGRAVOVHLVGEVQVYNSQGSVINDHIIIIIRQMLR RVTIISGDAELLPGELVETKFTENRNVVQSGHPASGRPOLMGTIKASLATESML SASFOETFLVTDAINAKSDSLIGIKENVIIGKLIIPACTGLSYRNIRNIRVEPTEEAK AAMYSAVGDDYDDISPPFGTSGQAVPLEDDYDGYPNQ"

4621..6402

/gene="SCO4655"

/note="Pfam match to entry PF00623 RNA_pol_A, RNA polymerase beta subunit, score 1264.40, E-Value 0"

misc_feature

misc_feature

RBS

gene

CDS

misc_feature

gene
polymerase alpha subunit, score 859.20, E-value 1.3e-254"
7703.8290
/gene="SCO4656"
/note="synonym: SCD40A.02"
7703.8290
/gene="SCO4656"
/note="SCD40A.02, possible integral membrane protein, len:
195 aa. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB81847.1"
/db_xref="GI:7288052"
/db_xref="UniProt/TREMBL:Q9L0K7"
/translation="MSDDKPFNNLLVSGVALLNLLIWGGAGSFAFLWDFAGRAS
NCPRIYPCPEGTYGLAGPVMALALALLAHLHREGVRGMPARAVPVLGAVLG
ARFLMSAYFFADDESRIYMAVVGVLGLPGALYGRFLRWGAAILLWGLDRGG
PVPASGPAQEGARNILLSLTGLVILGACGAGNALV"
complement(8307..9611)
/gene="SCO4657"
/note="synonym: SCD40A.03c"
complement(8307..9611)
/gene="SCO4657"
/note="SCD40A.03c, possible integral membrane protein,
len: 434 aa; similar to TR:Q9EMX4 (EMBL:AL445403)
Streptomyces coelicolor putative integral membrane protein
2SC134.05, 396 aa; fasta scores: opt: 678 Z-score: 732.5
E(): 3.6e-33; 7.923% identity in 414 aa overlap. Contains
possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB81848.1"
/db_xref="GI:7288053"
/db_xref="UniProt/TREMBL:Q9L0K6"
/translation="WTAMDADDTVDGPDGTERPAGSGAARSWRSLPALPPGA
RPLAPLALXGVTKLVGLVGFASLLTHAGDRGNRPFGGAHWDLVATWDGHWYIQ
VAEHGYPALVRLSDGLFTVQONSVAFFPLPALIRMSVESTGLGLVSGILVSVS
SFVAAGIYAVISLAGARAGTVAAGLWAPAGVWAVYSESFLVIAIAATCYAVM
KRWVTAGLALFTSPTSAVLGALVLAALVTLARPASREHVRGVPYAMLVAP
LGLLSYTAWGWSTGEATVETLQREGWAHPFDYGAIVTDVLRNLVGRGDYPPAFST
PDLLSLQLVLALPELLALMRLRRPPLVILVATLIVTLGTQOMFGNTRYLLPAPF

Query Match 13.1%; Score 242.4; DB 1; Length 292100;
Beat Local Similarity 50.4%; Pred. No. 1.2e-22;
Matches 736; Conservative 0; Mismatches 676; Indels 48; Gaps 4;

QY 215 CCAGCCCATCCGTGGCGAGCGCGGCGCGGATGCTGCTGCCCTCCGCGCGCGA 274
DB 288054 CGAGTCTTCGCGTGGCGCGCGCGCGCGCGATGAGCTTCCGAGGAGCGGTGGC 288113

QY 275 CCGGTGCGGTGCGTCCGAAGCGCGGCTCGGCGCTGTCTATCGCGAGGCGCGCCATGTCTA 334
DB 288114 GGCCTCGCGCGCGCGCTCCGTCTGAGACGCGCTGSCGCCAAGCGGACCCCGTCTA 288173

QY 335 CGGACTGACACCGGCTTCGGTCCCTTGGGAAACCGCTGTATCTCAGGTGAGATGTCG 394
DB 288174 CGGCGTGAGCACCGGCTTCGGCGCGCTCGGCGCGCACATGAGCCCGAGCTCGCGCG 288233

QY 395 AACCTGCAGGCAATCTTGTCATCATCTGCCAGCGCGCTGGGAGCGCGGTGCTTGACTG 454
DB 288234 CCGGCTCAGGCGCAACATCGTCCGCTGACGCGCGCGGATGGCGCGCGCTCGAGCG 288293

QY 455 GACGACGCGCGCGCATGGTTCCTGGCGCTGTGGTGTCAATCGCTCAGGGAGCCTCCGG 514
DB 288294 CGAGTGTTCGCGCACTGATGTTCTCGGCTGAAGACCGTCTGCTCCGCTCGCACCGG 288353

QY 515 TGCAGGAGGAGGACCATCGCTCGCTGATCGACTGCTCAATTCGAGCTCGCTCCGGC 574
DB 288354 CGTCCGCGCGGAGTTCGCGCGAGACCATGGCGGACGTGCTCAACGCGCGGATCACCCCGT 288413

QY 575 GGTTCACGCGCGCGCGGTGGGCGGTTCGGGTGACCTGACACCGCTTGGCATATGGT 634

DB 288414 GTGCAAGATAGGCTCTCTCGGCTGCTCGGTAACCTCGCCCGCTCTCCACTGTGGC 288473

QY 635 GCTCTGCTCCAGGCGCGGAGAGACTTCTTGAGACGGGACGGGACGCGCTTGACGGCGC 694

DB 288474 GCTGACCTCTATGCGAGGCGGACGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 288533

QY 695 AGAAGGCTCTCGGCGCGGACGCGTGCACCGCTCGATCTCTCCCATCGCGATGCACTGGC 754

DB 288534 CGAGCTGCTCGCCCGCGCACGCGGATCGCCCGCTCGAGCTGCGGAGAGAGGCGGCTCGC 288593

QY 755 GCTGTCAAACGGGACTCCGCGCATGACCGGATCGCGTGTGGTGAATGCTCAACGCTGCGC 814

DB 288594 CCTCTCAAACGGGACTGCTCGGATGCTGCTGATGCTGATGCTGATGCTGATGCTG 288653

QY 815 CCATCTCGGCAACTGGGCGGTGGCGTTGACGCGCTGCTGCTGCGGAATGCTTGAAGAGCGC 874

DB 288654 CACCTCTCAAGTCGGCGCGACATCACGCGCGCTGACCATGAGAGGCGCTGCTCGGCAC 288713

QY 875 GACGAGCATGGGCGCGGCGCATGTCGACTGCGGCGCGCATCCCGACAGAGAGCGC 934

DB 288714 CGACGCGTGTCTCGCCCGCGAGCTGCACGCCATCGCCCGCGGCGGCGGCGCGCAG 288773

QY 935 GCGAGCAGGCTGCGCGCGCGCTGGAGCGGCGGCGGCTGCTCGGCGCGCATATGCG 994

DB 288774 CGCGCCAAATGCGCGCGCTGCTCAAGGCTTCCGCGCTGACCGGCGCACACAGGACGA 288833

QY 995 CGAGCGGAGGCTCGACGCGCGCGCATGCGGACGAGCGGCGGCGGCGGAGTGCCTA 1054

DB 288834 GCGCCCGCGCTC-----CGAGTCCAAACGCACTTCCACGCGCGCGCTCGCTACGCTCTGGA 288857

QY 1055 CAGCTTGGCTGCGCTCGGAGGTTCTCGGCGCGGCTTCCACAGCTGCGATGCGATGA 1114

DB 288858 CTCGGTGGCTGTCTCGCGCGAGTTCGCGGCGGCGGCTGCGGCGCGCGCGCGCGG 288917

QY 1115 CCGGCTGTGACGATCGAGCTGAACGCGGTGACCGCAATCCGCTGTTTCCGCCCGATGG 1174

DB 288918 TCTGCTGCGCGAGCGGAGCTGGCGGCGGCGCTGACAAACCGGTGCTTCTCGCGACGG 288977

QY 1175 CAGCTGCGCGCTTGCACGCGGCGCAATTTTCATGGGCGAGCATGTGGCGCTGACGTCGGA 1234

DB 288978 ACGGT-----CGAGTCCAAACGCACTTCCACGCGCGCGCTCGCTACGCTCTGGA 289031

QY 1235 TGGCTGCGCGAGCGCTGACCGCTTCTGGCGGCGCTTGGCGGCGCGCGAGTTCGACGCT 1294

DB 289032 CTTCTCGCGTTCGCGCGCTCGCGCGCTCGGCTTCATCGCGAGCGCGCGCGCGCT 289091

QY 1295 GACAGATGAAGGCTGAACGCTGGGCTGCGCGGCTTCTCCACCGGCGCGCGCGGCT 1354

DB 289092 CTTGACAAAGAACCGGACCGGCGCTGCGCGCTTCTCGCGCTTCTCGCGCTTCTCGCG 289148

QY 1355 GAATTCGGCTTTCATGGGCGGCGACAGGTGACGGCGACCGGCTTCTGCGCGA---GATCG 1411

DB 289149 CGACTCGGCTGATGATGCGCCAGTACACGAGCGCGCGCTTGTGCGCGAGCTGAGCG 289208

QY 1412 AGCAACGGGCTGCTGCTGATTCATTCGATTCACGAAACCGCGCGCGCTTCAAGATGCTG 1471

DB 289209 GCTCGCGTACCGGCTGCGGCGGCTCCATCCGCTTCTCGCGATGACGAGGAGCCACGT 289268

QY 1472 CTCGTTTGGGACCATCGCGCGCGCTTTCGCGGAGAAAGATCGACGTTGCGGCGGAGAT 1531

DB 289269 GTGATGGCTGCTGCGGCGCGCGCAAGCTGCGACGCGCTGCAACCTTGGCGCGCT 289328

QY 1532 CTTTGGCATCTGCTCTCTGTTGCAAGCTGCGGAGTGGCGCTGCGGCGCGGCGCT 1591

DB 289329 GATCGCGTTCGAGCTGATCGCGCGCGCGGCGGATCCAGCTCCGCGAGGAGTACGCC 289388

QY 1592 AGACGGGCTGCTTCCCGCGGCGGAGAGCTGCTGCGGCGCGCTTCCGCGAGGATTCGCC 1651

DB 289389 GCGCGCGCGCTGCGAGGCGCTGCTGAGGCGCGTACGCGCGCGCGCTGTTGGAGGTTCCGG 289448

QY 1652 GCTTGAACGCGCGCGCGCT 1671

Db 289449 ACCGACCGCCACCTGGCGC 289468

RESULT 4

AP006840_33/c

WPCOMMENT

Sequence split into 36 fragments LOCUS AP006840 Accession AP006840

Fragment Name Begin End

AP006840_00 1 110000

AP006840_01 100001 210000

AP006840_02 200001 310000

AP006840_03 300001 410000

AP006840_04 400001 510000

AP006840_05 500001 610000

AP006840_06 600001 710000

AP006840_07 700001 810000

AP006840_08 800001 910000

AP006840_09 900001 1010000

AP006840_10 1000001 1110000

AP006840_11 1100001 1210000

AP006840_12 1200001 1310000

AP006840_13 1300001 1410000

AP006840_14 1400001 1510000

AP006840_15 1500001 1610000

AP006840_16 1600001 1710000

AP006840_17 1700001 1810000

AP006840_18 1800001 1910000

AP006840_19 1900001 2010000

AP006840_20 2000001 2110000

AP006840_21 2100001 2210000

AP006840_22 2200001 2310000

AP006840_23 2300001 2410000

AP006840_24 2400001 2510000

AP006840_25 2500001 2610000

AP006840_26 2600001 2710000

AP006840_27 2700001 2810000

AP006840_28 2800001 2910000

AP006840_29 2900001 3010000

AP006840_30 3000001 3110000

AP006840_31 3100001 3210000

AP006840_32 3200001 3310000

AP006840_33 3300001 3410000

AP006840_34 3400001 3510000

AP006840_35 3500001 3566135

Continuation (34 of 36) of AP006840 from base 3300001 (AP006840 Symbiobacterium the: morph

Query Match 13.0%; Score 241; DB 1; Length 110000;

Best Local Similarity 51.8%; Pred. No. 2.5e-22;

Matches 786; Conservative 0; Mismatches 665; Indels 66; Gaps 8;

Qy	183	GCGCTGAGCTGGATCGCCACATCGATCTGGACCGAGCCCATGCGGTGGCGAGCGCGGC	242
Db	104276	GCGGTAGAACTCGGTGACATTTAACCTTCCCGGAAGTCGTGGCTGCGCGAGCAGCGC	104217
Qy	243	GCGCGGATTTGCTTGGCCCTTCCGCGCGGACCGGTGCGGTGCGTCCGGAAGCGCGCTC	302
Db	104216	GCGCGCTGCTGCTGACCGCGGAGGTGGCGAGGTGGGTAGCCAGAGCGAGATGGTG	104157
Qy	303	GCGCGCTGTCATCGCGAGGCGCGCATGTCTVACGGAATGACAAACCGGCTTGGTCCCTT	362
Db	104156	GAACGGCTGTGGGAGCGCGCGCGCTTACCGCATCACACCGGTTTGGCAAGTTC	104097
Qy	363	GCGAACCGCTGATCTCAGGTGAGAAATGTCGACGCTGCGAGCGCAATCTTGTCATCAT	422
Db	104096	AGCGACGTCCCATATCGGACAGACAGACAGACATGACGCGCAATCTCTGATGAGC	104037
Qy	423	CTGGCCAGCGCGGTGGGACCGGTCTTGAATGACGACGCGCGCGCGCATGGTTCTGGCG	482
Db	104036	CACGCTTGCGCGTTCGAGAGCGCTCGCGCGGAAGTGGTGGCGGATGCTCTTCTG	103977
Qy	483	CGTCTGTGTGATCGTCTCAGGAGAGCTTCGGGTGCCAGCGAGGAGACCATCGCTCGCCTG	542
Db	103976	CGCGCCAGGCGCTCTCCCGGGGCACTCCGGCATCCGGGCGGAGACCTTGGAGATGCTC	103917

Qy	543	ATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGGTTCCTCCAGCGCGGACAGGTGGGCGC	602
Db	103916	GTGGGTTCTCAATCTCGGCTCACCCCGTGGTGCAGGAGAGGCTCCCTGGGCGCC	103857
Qy	603	TCGGGTGACCTGACACCGCTTGGCGCATATGGTGTCTCTCCAGGCGCGGAGAGATTTC	662
Db	103856	AGCGGACCTCGCGCGCTCGCCACATGAGCCCTCCGCTGATCGGGTTGGGTGAGGCC	103797
Qy	663	CTGGACCGGAGCGGAGCGGCTTGACGGCGCAGAGGGCTCCGGCGGAGCGCTGCAA	722
Db	103796	GTGTCAATGGCGA---GCGGCTGTCGCGCGCGAGAGCGCTGACAGCGGTCGGCTCCG	103740
Qy	723	CCGCTCGATCTCTCCCATCGCGATGCACTGCGCTGTGTCAACGAGACTTCGCCCATGACC	782
Db	103739	CCGCTGACGTGACGGCCAAGGAAGGGCTGGCTCTGATCAACGACCCAGCGCATGACC	103680
Qy	783	GGGATCGCGCTGGTGAATGCTCAACGCTCGCGCATCTCGGCAACTGGGCGGTGGCTTG	842
Db	103679	GCGCTGGGTCCTCGGCTCCACGATGCTCAGGTGCTGCTGAAAACCGCGGACATCGC	103620
Qy	843	ACGGCTGCTTGGGGAATGCTGAGAGGCGGAGACCGAGGATGGGCGCGGCACTGTCC	902
Db	103619	GCGGCCATGACCGGCGGCTGGGCGCGATCCCGCGGCTTGGGACCCCGGCTCGAG	103560
Qy	903	GACTCGCGCGCATCCCGACAGAGGACCGCGCAGCAGGCTGCGCGCCCGGTGGAC	962
Db	103559	GCCCTGCGGCTTCCACAGCGGCGAGCGGCTGTCAGGAACCTGCGCGCGGCT	103508
Qy	963	GGCAGCGCGGCTGGTGGTTCGCGCACGTCAATTCGCGAGCGGAGGCTCGACGCGCGGATATC	1022
Db	103507	-----TGACCGAGGGTTCGCGCTTACC-----ACC	103482
Qy	1023	GGGACGAGCGGAGGGGGGAGGATGCTCTACAGCTGCGCTGCGCTCCGAGGTTCTC	1082
Db	103481	CGACCGGGGAGATGCGCACCCAGGATCCGTACACCTTGCCTTCCGCGAGGTGCAC	103422
Qy	1083	GGCGCGGCTTCGACACGCTCGCATGACCGGTGTGACGATCGAGCTGAAGCGC	1142
Db	103421	GGCGCCAGCGGACCGGATCGAGCAGCTGGCGGAGGTGTGACTGGAGATGAGCGCC	103362
Qy	1143	GTGACCGACAATCCGGTGTTCGCCCGCATGGCAGCGTGCCTTGCACGCGGGCAAT	1202
Db	103361	GTACCGACAATCCCTGCTCTTCCCGGACGACGAGAG---GTATCTCCGCGGCAAC	103305
Qy	1203	TTGATGGCCAGATGTGGCGCTGACGTCGATGCGCTCCGACGCGGCTCACCGTTCTG	1262
Db	103304	TTCCAGGTCAGCGGTCGCGCTGGCGCTGGAATCTTGGCCATCGCCGTCGCGGAGCTC	103245
Qy	1263	GCGGGCTTCGCGAGCGCCAGATTGACAGTCTGACAGATGAAAGGCTGAACCGTGGGCTG	1322
Db	103244	GGGACATCGCCAGCGGCGCATCGAGGACTGGTCAACCCACAGCTTTCG---GGGCTG	103188
Qy	1323	CCCCCTTCTCCACCGGGGCCCCCGCGGTTGAAATTCGCGCTTCATGGGCGACAGGTG	1382
Db	103187	CCGGCTTCTCACCC---GCAACGGCGGGTGCATTCGGGGCTGATGATCAACCCAGTAC	103131
Qy	1383	ACGGGACCGCGCTCTGGCCGAG---ATCGGAGCCACGGGACCTGCTCGATCCATTCG	1439
Db	103130	ACGCGCGCTCGCTGGTGGAGCGAGAACAGGTGCTGCTCACCCGCGCAAGCTGGACTCC	103071
Qy	1440	ATCTCCACGAACCGCGCAATCAGGATGTGGTCTCGCTTGGGACCATCGCGCGCGCTC	1499
Db	103070	ATTCCGCTCAGCGCCAAACAGAGGAGCACGCTGTCGATGGGACACCGGCGCGCCGCAAG	103011
Qy	1500	TGCGCGGAGAGATCGACCGTTGGGCGGAGATCTTTCGATCTCTGCTCTCTGTTGCA	1559
Db	103010	GCGCGGAGGTGATCGCAACGCTCCGCGGCTGCTGGCCATCGAGCTGCTGCGCGGCC	102951
Qy	1560	CAAGCTCGGAGCTGCGCTCGGCGAGCGGCTTAGACGGGCTGCTCCCGCGGGGAGAG	1619
Db	102950	CAGGCTCTGGAGTTCTGGGCGCGGAGCGGCTCGCCCCCGCCACGCGCGCGCTACGC-	102892

ASLAPFWAHLILGGEVAVTQFCGAVEVGSJLHGLNVRVVEGRPTTWPSPDRSQOJHLDL
TVDDLDTAEHSAVRILGATREHQSPSPERTVLRDPAGHPFCLRA"
complement (11363). .12862)
/note="similar to Streptomyces coelicolor putative
rhannose kinase in NP_625114"
/codon_start=1
/transl_table=11
/product="probable sugar kinase"
/protein_id="AAZ23051.1"
/db_xref="GI:71068196"
/translation="MTTPPSALASTFAAVDLGATSGRVIVRGVRENLDLTEAHRG
PNTVRLPDGLRWDVALFOCALDGLREARSGVASVDWTWADHGLDADGALLG
HPHYRSDRTDAAARVHSLGPEELYRITGLQHLPTTVFOLAAEGTVQRAARTL
LLIPLLHMLTSGSAEETNASTGLPDARTCTWSODLLRRLLSLDPLGLPOLARGD
PAGTILLPHVAAYTGLGPRTPVTIVASHDTSAVAVPATPFDAYISCTGWSLAGLGL
DAPVITSESRANFTNBERGIDGTIRYLIRINMGWMLLEECRRTWDRNLPAHLSALLAH
AARAEPTASVLDPPAPVLPAGFDMPRIYDCARTQORVPSQGAIVRCVLESIALAH
RRALRQAALAGLRINDIRIHVQSGRRNDLLCOLTADATGLPVVAGTPEATGALGNILVO
ARATGLVDGLADMERLLIASTQHLRHYTPRGDGTAWHAAARLDLPAPGHR"
complement (13059). .15098)
/note="similar to Streptomyces avermitilis putative
sorbitol-6-phosphate 2-dehydrogenase in NP_828590"
/codon_start=1
/transl_table=11
/product="probable dehydrogenase"
/protein_id="AAZ23052.1"
/db_xref="GI:71068197"
/translation="MATHPEVAALLGRAHLGLSDPRNTVAGNSASAKIGVITDPTVGS
NTEMLVKGSGDITLTDEGLVALRLDRLRALQDVYVGREDEMVAFAFDHCLHGRC
GAAPSIDTAAKHELVADPHVLDHPSGLIACAADEGKLTABECYGAFAVWPVHRREG
QLGDI10VMEAGHPDVVGVVLGGHGI7AWGSGAQCEANALWMIRTAEEAFVLRBERG
PFGALEG7GALGAERERRAAALAPVIRSLASQDRPOVGHFTDSEVLDFDLAHDHP
HLAAGTSCDPFLTKRVPLVLDLPADALDEAVARLKHLEAEYEEVAAVYVRHAT
NAPSVMGADPAIVLIVGCMHSPGRDQKTARVAGEFYFNAINMGEAEAVSSYAPIE
EAEKFR1EYLEBEAKRMPKPKPLA1RVALTGTAGSGIGKAVARRLAEGACVVVA
DLNGDAAAABEGSGADKAAVAVTDTVSEEQ1AAAFREAVLAFGGVDLVVNNAGISV
KPSLETTAKRDMLQHD1IMARGSLFVREAA1VMTVQGLGSDIVYIVSKNAFATYGN
IAYSATKADQAOHVL1AALEBHGIRVNGVNPQVVGSGI7PAGCGAKRAATYCID
EEKLGEFYARRTLKRVLPHEVANA1FVALTGGELSHTTGLHPVDAGVAAAPLR"
complement (15190). .16359)
/note="similar to Streptomyces avermitilis putative sugar
isomerase in NP_828591"
/codon_start=1
/transl_table=11
/product="probable sugar isomerase"
/protein_id="AAZ23053.1"
/db_xref="GI:71068198"
/translation="MLMDDAVKAALKTKTQVTPSPMGYNSGTRFKVFAQGVGVPRTFW
EKLSDAAKHEFTGTPKVSLSHT1PMWRDVTDTALAAAYABERGLRGA1INSTVPODDY
RLG53HPDKDARRVAGHLECLSD1MDMADTGLD2KLWFDAGTNYTQGDQDWWARODRL
AAALAEVYERQORLLRLLBYKL7PEFAPY1TVDVDMGTSYAHCLKLPKPAQVVDRTGH
HAPCT5TEFTVAFLR8EKLGSYRFRFVADDMLVGSSDQFQLFRI1MHEVTKNGEL
APANAM1DQCHNIEAK1PAVIR5VMVQEATAKALLVDLDA1AAARQREGDVLAA
AALMDA1YTVRPLR1ELREEQGLAP1PLAAY1AASGWRK1EATRVGG5QAQWGA"
16364). .18278
/note="similar to probably glycosidase; reading frame is
interrupted; similar to Streptomyces narbonensis NbmF
encoded by GenBank Accession Number AAM88355"
18507. .19304
/note="similar to Streptomyces avermitilis hypothetical
protein in NP_822661"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAZ23054.1"
/db_xref="GI:71068199"
/translation="MRRGRCAAYIASNRGKPRHSSLLTPNIS1LNGLSVITIGISPDVA
DAFPD1ALIAV1TGLRGH6PEWATAAL2TLRQRLAAGTQWQADSDSDPR1SAWHVTY
RSFQTNPR1R3V2DALGRMWKGT1LR1NPADVA5YNAVSRH9LGPACGAFDLPHVTY
DVID1RYADGSSET7TFLP6GT2VBNPKPEGVYAD2DVL1TRHWNHDA1RTR1RVTEDST
HVA1PDLT1HATROGHLK1TAA1VELOSLA1PHAEOT1TVH1YAP1EACA1"
complement (15190). .16359)
/note="similar to Streptomyces avermitilis putative sugar
isomerase in NP_828591"
/codon_start=1
/transl_table=11
/product="probable sugar isomerase"
/protein_id="AAZ23053.1"
/db_xref="GI:71068198"
/translation="MLMDDAVKAALKTKTQVTPSPMGYNSGTRFKVFAQGVGVPRTFW
EKLSDAAKHEFTGTPKVSLSHT1PMWRDVTDTALAAAYABERGLRGA1INSTVPODDY
RLG53HPDKDARRVAGHLECLSD1MDMADTGLD2KLWFDAGTNYTQGDQDWWARODRL
AAALAEVYERQORLLRLLBYKL7PEFAPY1TVDVDMGTSYAHCLKLPKPAQVVDRTGH
HAPCT5TEFTVAFLR8EKLGSYRFRFVADDMLVGSSDQFQLFRI1MHEVTKNGEL
APANAM1DQCHNIEAK1PAVIR5VMVQEATAKALLVDLDA1AAARQREGDVLAA
AALMDA1YTVRPLR1ELREEQGLAP1PLAAY1AASGWRK1EATRVGG5QAQWGA"
16364). .18278
/note="similar to probably glycosidase; reading frame is
interrupted; similar to Streptomyces narbonensis NbmF
encoded by GenBank Accession Number AAM88355"
18507. .19304
/note="similar to Streptomyces avermitilis hypothetical
protein in NP_822661"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAZ23054.1"
/db_xref="GI:71068199"
/translation="MRRGRCAAYIASNRGKPRHSSLLTPNIS1LNGLSVITIGISPDVA
DAFPD1ALIAV1TGLRGH6PEWATAAL2TLRQRLAAGTQWQADSDSDPR1SAWHVTY
RSFQTNPR1R3V2DALGRMWKGT1LR1NPADVA5YNAVSRH9LGPACGAFDLPHVTY
DVID1RYADGSSET7TFLP6GT2VBNPKPEGVYAD2DVL1TRHWNHDA1RTR1RVTEDST
HVA1PDLT1HATROGHLK1TAA1VELOSLA1PHAEOT1TVH1YAP1EACA1"

Query Match

12.8%; Score 236.4; DB 1; Length 126933;

Best Local Similarity 51.0%; Pred. No. 9.3e-22; Matches 697; Conservative 0; Mismatches 626; Indels 45; Gaps 4;	
Qy	213 GACCAGGCCCATGCGTGGGAGCGGCGCGGATGTCTTGCCTCCGCGCGC 272
Db	4864 GACGAAGTGTGTCGTCGCGCGTCTCGGACCCCGTGTCTCGTCCGAAGCGCTG 4805
Qy	273 GACCGGTGCGTGCCTCGAAGCGCGCTCGCGCTGTCTCATCGCGAGGCGCGCCATGTC 332
Db	4804 AAGGAGATCCCGCAGCGCGGAGGTCGTACAGGCTCTCGCGACGACGCAAGCGCAC 4745
Qy	333 TAGCGACTGACACCGGCTTCGCTCCCTTGCAGAACCGGCTGATCTCAGGTGAGAAATGTC 392
Db	4744 TACGGCGTCTCCACCGGTTTCGGTGGCGTGGCCACCGCCACATCCGCCACCGAGCTGCGC 4685
Qy	393 CGAAGCGTGCAGGCCAATCTGTTCATCATCTTGGCCAGCGCGCTGGGACCGGTGCTTGAC 452
Db	4684 GCCAGCTCAGCGCGGTCTGGTGGCTGCCACGCGCGGCTTCGGACCCGAGGTGCGAG 4625
Qy	453 TGGACGACGGCGCGCCATGGTTCTTGGCGGCTCTGGTGTGATCGCTCAGGAGAGCTCC 512
Db	4624 CGCGAGTGTGTCGCGGCTCATGTCTGCTGGACTGTCCACGCTCGCCACCGGCCGTACC 4565
Qy	513 GGTGCCAGCAGGGGACCATCTGCTCGCGTGTACGACCTGTCTCAATTCGAGCTCGCTCCG 572
Db	4564 GGTGTCCGACCGACGAGACGGCGGAGGCTACGCGCGCTGTCTCAACCGCGGATCACTCCG 4505
Qy	573 GCCGTTCCAGCGCGCGACGCTGGTGGCGGCTCGGTGACTGACACGGCTTCGCGCATATG 632
Db	4504 GTCGTGACAGGTACCGGCTCGCTCGGTGTGTTTCGGGTGACCTCGCCCCCTTCGCGCACTGC 4445
Qy	633 GTGCTCTGCGCTCCAGGGCGCGGAGACTTCTCTGGAACGGGACGGGCTTGACGGC 692
Db	4444 GCGCTCGCGTCTATGGCGGAGGGGACGGTCCGACCGCGCGCTGCGCAGGCGCGGCG 4385
Qy	693 GCAGAAGGCTTCGGGCGCGGACGGCTGCAACCGCTCGATCTCTCCCATCGCGATGCACTG 752
Db	4384 GCGAGGCGCTCCCGGAGGCGGCGCATCACCCCGTCTGCTGAGGAGAAAGGAGGCGCTG 4325
Qy	753 GCGCTGTTCAAGGGACCTCGCCATACCGGGATCGCGCTGGTGAATGCTCAGCGCTGC 812
Db	4324 GCGCTCATCAAGGGACCGACGGCATGCTCGGATCTGGCGCTGGCGCGCCACGATCTG 4265
Qy	813 CGCATCTCGGCAACTGGGCGGTGGCGTTGACGGCCCTGTTGCGGAATGCTCTGAGAGGC 872
Db	4264 CGCGGCTCTCTCGCACGGCGACATCGCGCGCGCATGAGGTCGAGGGCCAGCTCGGC 4205
Qy	873 CGGACCGAGGCATGGGCCCGGCACTGTCCGACTCTGCGCCCGCATCCCGGACAGAAAGGAC 932
Db	4204 ACCGACCGGTCTTCGCCCGCAGACCTTCAGGCGCTCGCGCCGACCCCGGCGCAGGCGGAC 4145
Qy	933 GCGCGACGAGGCTGCGCGCGCGCTGGAGGCGGCGCGGGTGGTTCGCGCACGTCATT 992
Db	4144 AGCGCGCCCAATCTGGGGGCTTGTCTCGCGGCTCGCGATCTGGCCAGCCACAAGGCGC 4085
Qy	993 GCGGAGCGAGGCTCGACGCGGCGATATCGGGAACGAGGCGGGGCGAGGATGCC 1052
Db	4084 CCGGAGTGCAGGCGCT-----GAGGAGCGCC 4058
Qy	1053 TACAGCTCGGTGCGCTCCGAGAGTTCTCGGGGGGGGCTTCGACACGCTCGCATGGCAT 1112
Db	4057 TACTCGCTCCGTGACGCGCCAGAGTCCACGGCAACGCGCGGTGACACCTCTCTCCACGCG 3998
Qy	1113 GACCGGAGTGTGACGATCAGCTGAACGCGGTGACCGGAATCCCGGTGTTTCGCGCCGAT 1172
Db	3997 GAGCTGTTGGCGCGCCCGAGCTCGCTCCGCGCATCGCAACCCCGTCTGTCACGCGTGCAC 3938
Qy	1173 GGCAGGCTGCCGCCCTGCACGCGGGCAATTTTCATGGGCGAGCATGTGCGCTGACGTCC 1232
Db	3937 GCGCGGGTG-----GAGAGCAACGGCAACTTTCACGGCGGCTCCGCTCGCGCGGTGCTC 3884
Qy	1233 GATGCGCTCGCCACGCGCGTCAACCGTTCTGGCGGGGCTTCGCGAGCGCCAGATTGACGT 1292

Db 3883 GACTTCTCGCCATCTCGGTCCCGACGTCGCGTCGATCTCCGAGCGCCGACGACCGG 3824
Qy 1293 CTGACAGATGAAGCGTGAACCGTGGCTGCCCCCTTCTCCACCGGGGCGCCGCGG 1352
Db 3823 TTCCTGACCGCGCGCAATCGCGGCTCAACGCCCTTCTTCGGC---GAGACCCCGGA 3767
Qy 1353 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGGACCGGGCTCTCGGCCG---GATG 1409
Db 3766 GTGGACTCCGGCCACATGATCGCGCAGTACACCCAGCGCGCATCGTCTCGAGCTGAAG 3707
Qy 1410 CGAGCCAGGAGACCTGCTCATCTGATCTCCAGATCTCCAGAAAGCGCGCAATCAGGATGTG 1469
Db 3706 CGGCTCGCGCTGCGGGCTCCACGACTCGATCCCTTCCAGCGCCATCGCAGGAGGACAC 3647
Qy 1470 GTCTCGCTTGGGACCATCGCGCGCGCTCTGCGCGGAGAGATCGACCGTTGGCGCGAG 1529
Db 3646 GTCTCCATGGCTGGTCCGCGCGCGCAAGCTGCGCGCGCCCTCGACGGTCTCGGCGCG 3587
Qy 1530 ATCCTTGGATCTCTCGTCTCTGTCTTGTGCAAAAGCTGGGAGCTGCGC 1577
Db 3586 GTCGTGCGGTGGAACCTCTCACCGCGCGCGCGCCCTCGACCTGCGC 3539

RESULT 7
BA000030_41/c
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End
BA000030_00	1	110000
BA000030_01	100001	210000
BA000030_02	200001	310000
BA000030_03	300001	410000
BA000030_04	400001	510000
BA000030_05	500001	610000
BA000030_06	600001	710000
BA000030_07	700001	810000
BA000030_08	800001	910000
BA000030_09	900001	1010000
BA000030_10	1000001	1110000
BA000030_11	1100001	1210000
BA000030_12	1200001	1310000
BA000030_13	1300001	1410000
BA000030_14	1400001	1510000
BA000030_15	1500001	1610000
BA000030_16	1600001	1710000
BA000030_17	1700001	1810000
BA000030_18	1800001	1910000
BA000030_19	1900001	2010000
BA000030_20	2000001	2110000
BA000030_21	2100001	2210000
BA000030_22	2200001	2310000
BA000030_23	2300001	2410000
BA000030_24	2400001	2510000
BA000030_25	2500001	2610000
BA000030_26	2600001	2710000
BA000030_27	2700001	2810000
BA000030_28	2800001	2910000
BA000030_29	2900001	3010000
BA000030_30	3000001	3110000
BA000030_31	3100001	3210000
BA000030_32	3200001	3310000
BA000030_33	3300001	3410000
BA000030_34	3400001	3510000
BA000030_35	3500001	3610000
BA000030_36	3600001	3710000
BA000030_37	3700001	3810000
BA000030_38	3800001	3910000
BA000030_39	3900001	4010000
BA000030_40	4000001	4110000
BA000030_41	4100001	4210000
BA000030_42	4200001	4310000
BA000030_43	4300001	4410000
BA000030_44	4400001	4510000

BA000030_45	4500001	4610000
BA000030_46	4600001	4710000
BA000030_47	4700001	4810000
BA000030_48	4800001	4910000
BA000030_49	4900001	5010000
BA000030_50	5000001	5110000
BA000030_51	5100001	5210000
BA000030_52	5200001	5310000
BA000030_53	5300001	5410000
BA000030_54	5400001	5510000
BA000030_55	5500001	5610000
BA000030_56	5600001	5710000
BA000030_57	5700001	5810000
BA000030_58	5800001	5910000
BA000030_59	5900001	6010000
BA000030_60	6000001	6110000
BA000030_61	6100001	6210000
BA000030_62	6200001	6310000
BA000030_63	6300001	6410000
BA000030_64	6400001	6510000
BA000030_65	6500001	6610000
BA000030_66	6600001	6710000
BA000030_67	6700001	6810000
BA000030_68	6800001	6910000
BA000030_69	6900001	7010000
BA000030_70	7000001	7110000
BA000030_71	7100001	7210000
BA000030_72	7200001	7310000
BA000030_73	7300001	7410000
BA000030_74	7400001	7510000
BA000030_75	7500001	7610000
BA000030_76	7600001	7710000
BA000030_77	7700001	7810000
BA000030_78	7800001	7910000
BA000030_79	7900001	8010000
BA000030_80	8000001	8110000
BA000030_81	8100001	8210000
BA000030_82	8200001	8310000
BA000030_83	8300001	8410000
BA000030_84	8400001	8510000
BA000030_85	8500001	8610000
BA000030_86	8600001	8710000
BA000030_87	8700001	8810000
BA000030_88	8800001	8910000
BA000030_89	8900001	9010000
BA000030_90	9000001	9025608

Continuation (42 of 91) of BA000030 from base 4100001 (BA000030 Streptomyces ave

Query Match	12.6%;	Score 232.8;	DB 1;	Length 110000;	
Best Local Similarity	50.0%;	Pred. No. 2.8e-21;			
Matches 730;	Conservative	0;	Mismatches 682;	Indels 48;	Gaps 4;

Qy	215	CGAGGCCCATGCGTGGCGAGCGCGCGCGGATGTCCTTGGCCCTCGGGCGCGGA	274
Db	39677	CGACGTTCTCGCGTGGCGCGTGGCGCGCGCGATCGAGCTCTCCGGGAGGCGGTGAC	39618
Qy	275	CCGGTCCGTGCGTCCGAAGCGCGCTGGCGCTGTCATCCGCGAGGCGCGCATGTCTA	334
Db	39617	CGCCCTCGCGCGCGCGGCGATCGTGGACGCGCTGGCGGCAAGCCGAGCCCGTGTA	39558
Qy	335	CGGACTGACAAACCGGTTCCGTTCCGAACCGCTGATCTCAGGTGAGAAATGTCG	394
Db	39557	CGGGGTGAGCACCGGTTTCGGGGCCCTTGGGACCCCGGCACATCAGCCAGGAGCTCGGGC	39498
Qy	395	AACGTCGAGGCCAATCTTGTTCATCATCTGGCCAGCGCGGTGGACCGGTGCTTGACTG	454
Db	39497	GCAGCTCCAGCGCAACATCTGCTCCGCTCGCACGCGCGCGCATGGGTCCGAGGGTGGAGCG	39438
Qy	455	GACGACGGCGCGCCCATGGTTCTGGCGGCTCTGGTGTGATCGCTCAGGAGGACCTCCGG	514
Db	39437	GGAGGTCTGTACGGGCGGCTGATGTTCTTCGCACTGAAAAACCGTCTGCTCGGGCCACACCGG	39378
Qy	515	TGCCAGCGAGGGGACCATCGCTCGCCTGATCGACCTGTCTCAATTCGAGACTCGCTCCGGC	574

Db 39377 CGTACGGCCGGAAGTTCGGCGACACGATGCGCCGACATCTCAACGCGGGATCACCCCGGT 39318
QY 575 CATTCCAGCCCGGACCGGTGGGCGCTCGGGTGAACCTGACACCGCTTGGCATATGTT 634
Db 39317 CGTGCACGAGTACGGCTCGCTCGGTCTCGCGGACCTCGCCCGCTCTCCACTGGGC 39258
QY 635 GCTCTGCTCCAGGGCGGGGAGACTTCTGACACGGGACGGGACGGCGCTTGACGGGCG 694
Db 39257 CTTGACGTGATGGGCGAGGGCGACCGCGAGGGCCCGGACCGCATCGTTCGTCGGCCGG 39198
QY 695 AGAAGGGTCCGGCGCGGACGGCTGCAACCGCTCGATCTCTCCATCGCGATGCACTGCG 754
Db 39197 TGAACCTCTCGCGAGCACGGCATACGCGGTGAGCTGCGCGAGAGGGGCTTGGC 39138
QY 755 GCTGTGTAACGGGACCTCGCCCATGACCGGATCGGCTGTGTAATGCTCAACGCTTGGCG 814
Db 39137 CTTCTCAACGACCGACCGGATGCTCGGATGCTGCTGATGCTGCTGCTGCTGCTGCTGCA 39078
QY 815 CCATCTCGCAACTGGGCGGTGGCTTGAACGGCTCTGTCGGGATGCTGAGAGGCGG 874
Db 39077 CGCGCTTACAACTGCGGACATACGCGCGCCCTCTCCCTCGAGGCATCTCTCGGCAC 39018
QY 875 GACCGAGGATGGGCGCGGACCTGTCGACCTCGCGCGCATCCCGGACAGGACGCG 934
Db 39017 CGACAGGTGCTCGCCCGCAACTGCAACCCATCGCCCGACCCCGGTTCAGGGCGCTTC 38958
QY 935 CGACGAGGTGCGCGCGCGCTGAGCGGACGCGCGGGTGTTCGCGACGTCATTGCG 994
Db 38957 GSCCGCCAACTGCTGGCGCTACTGGCGGTTCCGAGC----- 38920
QY 995 CGAGCGGAGGTCGACGCGCGGATATCGGACGAGGCGCGGAGGCGGGGAGGATGCTTA 1054
Db 38919 -----TCACGGCGCACACAGGACGACGCGCGCGCTCCAGGACGCGTA 38874
QY 1055 CAGCTGCGCTCGCGCTCGCGAGTTCTCGGGCGGCTTCGACACGCTCGCATGSCATGA 1114
Db 38873 CTCGTGCTGCTCGCCCGCAGGTGCGCGCGCGCGCGCGGACACCTCTGCGCCACGCGCG 38814
QY 1115 CCGGTGCTGACGATCGAGCTGAACCGGTGACCGCATCCGTTGTTTCGCGCGCGATGG 1174
Db 38813 CTTCTGCGGAGCGGAGCTGGCGTTCGGGTGACAAACCGGTGCTGCTCCCGA--- 38757
QY 1175 CAGCTGCGCGCTCGACGGGGGCAATTCTATGGGCGAGCATGTGGGCTGACGTCGGA 1234
Db 38756 ---CGGGCGCTGAGTCCAAACGGCACTTCACGCGTGCCTGCTGCGGTACGCTCTCGA 38700
QY 1235 TGCGTTCGCGCGCGCTCACCGTTCTGGCGGCTTGGGAGCGGACGATTTGACGCTCT 1294
Db 38699 CTTCTCGGATCGCGCGCGGACCTCGGTTCCATCGCGGCGCGCGACCGACCGCGCT 38640
QY 1295 GACAGATGAAGGCTGAACCGTGGGCTGCGCCCTTCTTCCACCGGGGCGCGCGGGTT 1354
Db 38639 TCTCGACAGAACCCTCGCACGGGCTGCGCGGTTCT---CGCGGACGACGCGGGCT 38593
QY 1355 GAATTCGGGCTTATGCGGCGCACAGGTGACGCGGACCGGCTCTTGGCGGAGAT---GGG 1411
Db 38582 CGACTCGGCGCTGATGATCGCCAGTACACGAGCGCGGCTTGTGTCAGCGAGATGAAGCG 38523
QY 1412 AGCCACGGGACCTGCTCGATTCGATTCGATTCGACGACGCGCGCAATCAGGATGTGTT 1471
Db 38522 GCTCGCGTTCGCGGCTCGGAGATCGATCCGTCCTCGGCGATGTCAGGAGGACCACT 38463
QY 1472 CTCGTTCGGACCATTCGCGCGCGCTCTGCGCGGAGAGATCGACCGTGGGCGGAGAT 1531
Db 38462 CTCGATGGGGTGGTGGCGCGCCCGCAAGCTCGTACGAGGATCGAACCTCACGCGCAT 38403
QY 1532 CTTTCGATCTCTGCTCTGCTTTGCAAAAGCTCGGAGTGTGCTCGGCGAGCGGCT 1591
Db 38402 CGTGGCGATCGAGCTCTACGCGCGCACCGCGGCCATCGAACTTCGCGAGGGTCTGACCCC 38343
QY 1592 AGACGGGTGTTCTCCGCGGGGGAAGAGCTGTGTGACGGCCCTGCGCGAGCATTTCCGCG 1651

Db 38342 CGCCCCGGGCTCGCAGGCGGTTCATCAGCGCCGTCCGCAAGGCGGGCTCGAAGGCCCAGG 38283
QY 1652 GCTTGAGACGACACCGGCC 1671
Db 38282 ACCGACCGCTTCTTGGGCC 38263

RESULT 8
CP000010_06
WPCOMMENT

Sequence split into 36 fragments LOCUS CP000010 Accession CP000010

Fragment Name	Begin	End
CP000010_00	1	110000
CP000010_01	100001	210000
CP000010_02	200001	310000
CP000010_03	300001	410000
CP000010_04	400001	510000
CP000010_05	500001	610000
CP000010_06	600001	710000
CP000010_07	700001	810000
CP000010_08	800001	910000
CP000010_09	900001	1010000
CP000010_10	1000001	1110000
CP000010_11	1100001	1210000
CP000010_12	1200001	1310000
CP000010_13	1300001	1410000
CP000010_14	1400001	1510000
CP000010_15	1500001	1610000
CP000010_16	1600001	1710000
CP000010_17	1700001	1810000
CP000010_18	1800001	1910000
CP000010_19	1900001	2010000
CP000010_20	2000001	2110000
CP000010_21	2100001	2210000
CP000010_22	2200001	2310000
CP000010_23	2300001	2410000
CP000010_24	2400001	2510000
CP000010_25	2500001	2610000
CP000010_26	2600001	2710000
CP000010_27	2700001	2810000
CP000010_28	2800001	2910000
CP000010_29	2900001	3010000
CP000010_30	3000001	3110000
CP000010_31	3100001	3210000
CP000010_32	3200001	3310000
CP000010_33	3300001	3410000
CP000010_34	3400001	3510000
CP000010_35	3500001	3510148

Continuation (7 of 36) of CP000010 from base 600001 (CP000010 Burkholderia mallei ATCC 21

Query Match 12.5%; Score 231.6; DB 1; Length 110000;
Best Local Similarity 50.9%; Pred. No. 4e-21;
Matches 753; Conservative 0; Mismatches 664; Indels 63; Gaps 6;

QY 229 TGGCGAGCGGCGCGCGGATTTGCTTGGCCCTCCCGCGCGGACCGGTGCGGTGCGT 288
Db 71181 TCGCCCGGAGACGTGCGATCGCGCTCGATCCCGAGCTTCGCCGCGATCGACCGGG 71240
QY 289 CCGAAGCGCGCTCGGCTGTCTTCGCGAGGCGCGCATGTCTACGAGCTGCAACACG 348
Db 71241 GCGCGAGCGCTGCCGACATCGCGGAGAGGGGAGCGCGGTACGCGATCAACACG 71300
QY 349 GCTTCGGTCCCTTCGCAACCGCTGTATCTCAGGTGAGAATGTCCGAACGCTGCGAGGCA 408
Db 71301 GCTTCGGCGCTTCGAGACGCAATTCGCGACGACGAGCTCGAGCTGTGCGAGAAGA 71360
QY 409 ATCTTGTCCATCATCTGCGCAGCGGTGGACCGGTGCTTGACTGGAGCAGCGCGCGG 468
Db 71361 ACCTGCTGCTTCGACACGCGGTGGGCGTCCGCGAGCCGATGCGCGCCCGCTGTCGCGC 71420
QY 469 CCATGGTTCTGCGCGCTGTGGTGTGATCGCTCAGGGAGCCCTCGGTGTCCAGCGAGGGA 528
Db 71421 TGTGATGCGCGCTCAGCTCTCGAGCCTCGCGCGCGCACTCGGGCATTCGTCCGCTCG 71480

QY 409 ATCTTGTCCATCATCTGGCCAGCGGCTGGGACCGGTGCTTGACTGAGCGAGCGGCGCG 468
Db 33775 ACCTGTGTGTCGACAGCGGCTGGGCTCGGAGCGCGATGGCGGCCCGCTGTGCGCC 33716
QY 469 CCATGTTCTGGCGGCTCTGGTGTGATCGCTCAGGAGCCTCCGGTGCACGAGGAGGGA 528
Db 33715 TGTGATGGCGCTCAAGCTCTGAGCCTCGGCGGCGCCACTCGGCGCATTCGTCGCTCG 33656
QY 529 CCATCGCTCGCTGATCGACTGCTCAATTCGAGCTCGCTCGGCGCGTTCGACGCGCG 588
Db 33655 TGATGAGCGCTCGTCAAGCTGTTCAACCGGAGCGTGTGCGCGTCAATTCGCGTCAAG 33596
QY 589 GCACGCTGGCGCGCTCGGCTGACCTGACACACCGCTTGGCATATGTGTGCTCTCCCTCCAG 648
Db 33595 GCTCGGTGGCGCGCTCGGCGACTTGGCGCGCTCGGCGACATGTGCGCGCTGCTGCTG 33536
QY 649 GCCGGGAGACTTCTTGACCGGAGCGGAGCGGCTTGAACGCGCGAGAGGGCTCCGCG 708
Db 33535 GCATCGGCGACGTGTTTCATCCGCGCGA---GCGCGGAGCGCGCGAAGGGCTGCGTG 33479
QY 709 GCGGAGGCTGAACCGCTCGATCTCTCCATCGGATGCACTGGCGGTGCTGTAACGGGA 768
Db 33478 TCGCGGCGCTCGCGCGCTACGCTCGAAGCGAAGGAGGCGCTCGCTGCTGTAACGGGA 33419
QY 769 CCTCGCGCATGACCGGATCGGCTGCTGATGCTCAGCGCTCGCGCATCTCGGCAACT 828
Db 33418 CCGAGCGTGAACCGCTCGCGCTCGACACCTGTTGCGATCGAGGACCTGTACCGGA 33359
QY 829 GGGCGGTGGCTTGACGGCCCTGCTTGGCGAATGCTGAGAGCGCGAGCGGATGGG 888
Db 33358 CGGCGCTCGTTCGGCGCGCTGTCGTCGACGCGCGCGGCTCGGTGAAGCGGTTGCG 33299
QY 889 CCGCGGCACTGTCGACCTCGGCGCGATCCCGGACAGAGAGCGCGGAGCGGCTGCG 948
Db 33298 AGCGCGCATCCAGAGTTCGCGCGCATCGCGCGAGATCGACGCGCGCGCGCTACC 33239
QY 949 GCGCCCGGTGACCGGAGCGCGGCTGTCGCGACGTCTATTCGCGAGCGGAGGCTCG 1008
Db 33238 GGTGCTGCTCGACGGCTCGCGGATCAACGTGTGCGACCGC----- 33198
QY 1009 ACGCGCGGATATCGGACGAGCGCGAGCGGCGGCGAGTCCCTACAGCTGCGCTGCG 1068
Db 33197 -----GATTGCGACAAGTTCAGGACCGCTGACAGCCGTGCGCTGCGC 33158
QY 1069 CTCGCGAGTTCTCGGCGCGGCTTCGACAGCTCGCATGGCATGCGGCTGCTGACGA 1128
Db 33157 AGCGCGAGTGTATGGCGGCTGCTCGACAGATCGGCGACGCGCGGCTGCTGCTCA 33098
QY 1129 TCGAGCTGAACGCGGTGACGCAATCCGCTGTTTCGCGCGGATGGGAGCGGTGCGCGCC 1188
Db 33097 TCGAGCGAAGCGGCTATCGGACAAACCGCTGATCTTCCGAGACACGGGCGAG---GTGC 33041
QY 1189 TGCAGCGGGCAATTTTCATGSCCAGCATGTGGCGCTGACGCTGCGATGCGCTCGCAAG 1248
Db 33040 TGTGCGGGCGCAATTTTCACGCGGAGCGGCTGCGCTTGGCGCGCAATCTCGCGATCG 32981
QY 1249 CGGTACCGCTTCTGGCGGCTTTCGCGAGCGGCAATTCGAGTTCGATGAGATGAAGGC 1308
Db 32980 CCGCGCGGAGATCGGCGCTTCGCGAGCGGCGCATCGGCTGTTGATCGAGCGGAGCG 32921
QY 1309 TGAACGTTGGGTGCGCCCTTCTCTCAACGCGGCGCGCGGTTGAAATTCGCGCTTCA 1368
Db 32920 TCTCGGCGCTGCGGCTTCTCTGTTGAAGAGCGGCGG-----TGAACCTGGGCTTCA 32867
QY 1369 TGGGCGCAGGTGACGCGACCGGCTCTCGGCG---AGATGCGAGCGGAGACCTG 1425
Db 32866 TGATCGGCGACGTGACGCGCGCGCTCGCTCGGCTCGGAAACAAAGAGCTTCGCGCATCCGG 32807
QY 1426 CCTCGATCCATTCGATCTCCAGAAACGCGCGCAATCAGGATGTGCTCTCGTTGGGAGCA 1485
Db 32806 CGTCCGTGATTCGCTGCGGAGCTCGGCGNACGAGAGCAACGCTGTGATGGGAGCGT 32747
QY 1486 TCGCCGCGCGCTCTGCGCGGAGAGATCGACCGGTTGGGCGGAGATCTTTCGATCTCG 1545

Db 32746 TCGCGCGCGCAAGCTCGGAGCATCGGAGAAAGTTCGCGAATCTCTCGCGATCGAGC 32687
QY 1546 CTCTCTGTTCGACAAGCTCGGAGCTCGCTCGGCGAGCGGCTTAGACGGGTGTCTC 1605
Db 32686 TGCTCGCGCGCGCAAGCGCTCGACCTGC-----GCGCGCGGACGCAACGAGCC 32636
QY 1606 CCGCGGGAAGAGCTGTGAGGCGCTCGGCGAGAGTTCGCGCGCTTGAGACGGACC 1665
Db 32635 CGGCGCTGAGACGCGATGAAGACGATTCGCGCGGAGCGTTCGCACTACGATCTCGACC 32576
QY 1666 GCGCCCTCGGACAGGAAATTCGCGCTTGTACGCACT 1705
Db 32575 ACTACTTCGCGCCGACATCGGCTGCTCGCGCGCGGT 32536

RESULT 10
CP000031.23
WPCOMMENT

Sequence split into 41 fragments LOCUS CP000031 Accession CP000031

Fragment Name	Begin	End
CP000031_00	1	110000
CP000031_01	100001	210000
CP000031_02	200001	310000
CP000031_03	300001	410000
CP000031_04	400001	510000
CP000031_05	500001	610000
CP000031_06	600001	710000
CP000031_07	700001	810000
CP000031_08	800001	910000
CP000031_09	900001	1010000
CP000031_10	1000001	1110000
CP000031_11	1100001	1210000
CP000031_12	1200001	1310000
CP000031_13	1300001	1410000
CP000031_14	1400001	1510000
CP000031_15	1500001	1610000
CP000031_16	1600001	1710000
CP000031_17	1700001	1810000
CP000031_18	1800001	1910000
CP000031_19	1900001	2010000
CP000031_20	2000001	2110000
CP000031_21	2100001	2210000
CP000031_22	2200001	2310000
CP000031_23	2300001	2410000
CP000031_24	2400001	2510000
CP000031_25	2500001	2610000
CP000031_26	2600001	2710000
CP000031_27	2700001	2810000
CP000031_28	2800001	2910000
CP000031_29	2900001	3010000
CP000031_30	3000001	3110000
CP000031_31	3100001	3210000
CP000031_32	3200001	3310000
CP000031_33	3300001	3410000
CP000031_34	3400001	3510000
CP000031_35	3500001	3610000
CP000031_36	3600001	3710000
CP000031_37	3700001	3810000
CP000031_38	3800001	3910000
CP000031_39	3900001	4010000
CP000031_40	4000001	4109442

Continuation (24 of 41) of CP000031 from base 2300001 (CP000031 Silicibacter pomeroyi DS)

Query Match 12.3%; Score 227.2; DB 1; Length 110000;
Best Local Similarity 50.5%; Pred. No. 1.5e-20;
Matches 723; Conservative 0; Mismatches 658; Indels 51; Gaps 5;
QY 255 CTTGCCCCCTCGCGGCGCGGACCGGTCGCTCCGAGCGCGGCTCGGCTGCTCATC 314
Db 29285 CTTGACCGGCTCGCGCCCGGCTCGAGGCGGCGCGCATCGCGCGCGCG 29344
QY 315 CGGAGGCGGCCCATGTCTACGAGTGAACACCGGCTTCGTTCCCTTGGGAACCGCCTG 374

Db 29345 GCGGGCGATGTGCGGGTCTATGGGTCACACACGGGCTTTGGCAAGCTGGCCAGCTCAAG 29404
Qy 375 ATCTCAGGTGAGATGTCCGAACGCTGCAGGCCAATCTTGTCCATCATCTGGCCAGCGGC 434
Db 29405 GTCCGGCGGAGGACACCGCCACCTTGACGAACCTGATCTCTGTCCCATGTCTGGCG 29464
Qy 435 GTGGGACCGGTGTGTGACTGACACCGCGCGCCCATGTGTCTTGGGCGCTCTGGTGTCTG 494
Db 29465 GTCCGGCGCCGATCCCGCGCGCATGGCCGGCTGATGTGTCTTAACTGTCTGTCTG 29524
Qy 495 ATCGCTCAGGAGCTCTCGGTGCGAGGAGGAGACCATGTCTCCCTGTGATGACCTGCTC 554
Db 29525 CTGGGGGTGGCGCTCGGGCGTGTGGGTGGAGCTGATCACTGTCTTCAGGAGATGCTG 29584
Qy 555 AATTCCGAGCTCGTCCGGCGGTCCAGCCGCGGACGGTGGCGGTGGGTGACCTG 614
Db 29585 GCGCGCATGTCAACCCGGTAAATCCCGGTGAGGGCTCAGTCGGGGGCTCGGGGATCTG 29644
Qy 615 ACACCGCTTGGCATATGTGTCTCTGCTCCAGCGCGCGGAGACTTCTTGGACCGGAC 674
Db 29645 GCGCACTGGCCATATGACGGCGTGTATCATCGCGCAGCGGAGCGGATATC---AG 29701
Qy 675 GGGACGGCTTGTGACGGCGAGAGGGTCTCGCGCGGACGGCTGTGCAACCGCTCGATCTC 734
Db 29702 GGCCAGCGCTCTCCCGCGCGCGAGCGCTGGCGCGCGCTGGCTTGAACCCGATTCGCGTG 29761
Qy 735 TCCCATCGCATGCACTGGCGCTGTCTCAACGGGACCTCCGCCATGACCGGGATCGCGCTG 794
Db 29762 GGGCCCCAAGGAGGGGTGGCTTTATCAACGGCACCCAGTTCCTGACGCGATCTTCGCGCTG 29821
Qy 795 GTGAATGTCTACCGCTCTCGCGCATCTCGGCACTCGGCGGTGGGTGACGGCCCTGCTT 854
Db 29822 GCGGCGCTGTTCGGCGCTGGCGCGCGGACCTCTCTGTGTGTGAACGGGCACTGTCTG 29881
Qy 855 GCGGAATGTCTGAGAGCCGAGACCGAGGCAATGGCCCGCGGCACTGTCCGACCTGCGCGCG 914
Db 29882 ACCGATGCGATCATGGGTGCACCGCACCCCTGCAACCCGAAATCCACGCCCTGCGCGCG 29941
Qy 915 CATCCCGACAGAGGAGCGCGACGAGGCTGCGCGCCCGCTGGAGCGGACGGCGCGG 974
Db 29942 CATCGCGCCAGATCGACGCGCGCGCGGATGCGCGCTGTCTTGACGGGTCCGAGATC 30001
Qy 975 GTGCTCGGCGACGTCATTTCGCGAGCGAGGCTCGACGCGCGCATATCGGACGAGCGG 1034
Db 30002 CGCAAGCACCGGAGGCGACACGG----- 30030
Qy 1035 GAGCGGGGAGGATGCTACAGCTCGCTGCTGCTCGAGGTTCTCGGGGGGGCTTC 1094
Db 30031 -----GGTGCAGGATCCCTATTGCATCCGCTGCCAGCGGAGTGACGGGTGTGCCATG 30085
Qy 1095 GACACGTCGATGCGATGACCGGGTGTGACGATCGAGTGAACGCGGTGACCGACAAT 1154
Db 30086 GACGTGTGCGCAGGCGCGCCAGACCTGGAGATCGAGGCAATGCGCGCAACCGACAAC 30145
Qy 1155 CCGGTGTTTCGCGCGGATGGCAGCGTGCCTCGACGCGGGCAATTTTCATGGGCGAG 1214
Db 30146 CCGTGTGTGGCGGAGGCGGACATGATCG---TCTCTGGCGCAATTTTCATGCCGAG 30202
Qy 1215 CATGTGGCGTGAAGTCCGATGCGCTCGCCACGCGCTCAACGTTCTTGGCGGGCTTTCG 1274
Db 30203 CCGGTGGCTTTGCGCGGATCTGATCGACTGSCCTGTCCGAGATCGGCGCATCGCG 30262
Qy 1275 GAGCGCAGATGTGACGTCTGACAGATGAAGGCTGAACGCTGGGCTGCCCGCTTCCTC 1334
Db 30263 CAACCGCGGGTGGCGCTGATGGTGGATCCGACGCTGAGTTTCGATCTGCGCGCTTCCTG 30322
Qy 1335 CACCGGGCCCCCGGTTGAAATTCGGCTTCATGGCGGCACAGGTGACGGGACCGCG 1394
Db 30323 ACCCGGAGC---CGGGGCTGAATTCGGTCTGATGATGCGCGGAGTCAACCGCGCGG 30379
Qy 1395 CTCCTGCGCAGATGCGAGCACCGGACCTGCTCTGATTCATTCGATCTTCCAGAACGCC 1454

Db 30380 CTGATGAGCGAGAACAGCAATCTGGCAATCTCGGTGTACCGACAGACCCCCACACAG 30439
Qy 1455 GCCAATCAGGATGTGGTCTCGCTTGGGACCATCGCGCGCGCTCTTCCCGGAGAGATC 1514
Db 30440 GCCAATCAGAGG-----ATCAGGTATCATGCGCGGATGGGGGGTTCGGCTGGC 30493
Qy 1515 GACCGTTGGCGGAGATCTTGGCATCTCTCGCTCTTGTCTTGCACAAGCTGGGAGCTG 1574
Db 30494 CGGATGCTCGAGAACCTTGCAATTACTCTGGCGTTCGAGCTGCTTTCGCGCGCGAGGG 30553
Qy 1575 CGTGGCGGACGGCTTAGACGGGGTGTCTCCCGCGGGGAGAGCTGTGTGACGGCCCTG 1634
Db 30554 ATCGAGTTTCGCGCGCCCTTTCGACGACGACCGCGCTGCAAGCTGCGGTGCGCGCTG 30613
Qy 1635 CGCGAGCAGTTCCGCGCTTGGAGCGACCGCGCCCTGGGACAGGAAATTG 1686
Db 30614 CGCGCAGGTGCCCGGCTGGAGAGAACCGTTATATGCGCGCCGAGATCG 30665

RESULT 11
AP006618_13/c
WPCOMMENT
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618

Fragment Name	Begin	End
AP006618_00	1	110000
AP006618_01	100001	210000
AP006618_02	200001	310000
AP006618_03	300001	410000
AP006618_04	400001	510000
AP006618_05	500001	610000
AP006618_06	600001	710000
AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000

AP006618_49	4900001	5010000	
AP006618_50	5000001	5110000	
AP006618_51	5100001	5210000	
AP006618_52	5200001	5310000	
AP006618_53	5300001	5410000	
AP006618_54	5400001	5510000	
AP006618_55	5500001	5610000	
AP006618_56	5600001	5710000	
AP006618_57	5700001	5810000	
AP006618_58	5800001	5910000	
AP006618_59	5900001	6010000	
AP006618_60	6000001	6021225	
Continuation (14 of 61) of AP006618 from base 1300001 (AP006618 Nocardia farcinica fPM 1)			
Query Match 12.2%; Score 226.2; DB 1; Length 110000;			
Best Local Similarity 50.16%; Mismatches 623; Indels 45; Gaps 4;			
Matches 685; Conservative 0;			
Qy	228	GTGGCGAGCGCGCGCGGATTTCTTCTTCCCTCCCGCGCGGACCGGTGCGTGC	287
Db	55487	GTGGCGAGGAGCGGCGCGGCTGTCCGACGAGGCGGTGGCGCGATGGCCGAC	65428
Qy	288	TCCGAAGCGCGGCTGGCGCTGTATCCGCGAGGCGGCCATGTATCGACTGACAA	347
Db	65427	AGTCGAAGCGCATCGAGGCGTGGCTGCCGACCGAAGCCGCTCTACGGTGTCTCG	65368
Qy	348	GGCTTCGCTCCCTTGGCAACCGCTGATCTCAGTGAGATGTCCGAAACGCTGAG	407
Db	65367	GGGTTCGGCGGCTGGCGGTGGCCATATCCCGCTCGAGCTCCGCAAGCAGTTC	65308
Qy	408	AATCTTGTCCATCTCTGCGCAGGCGGTGGACCGGTGCTTGAATGACGACGCG	467
Db	65307	AGTCTGTGCGCTGCGACCGCGGCGGTCTGGGCCCGGAGTTCGAACGCGAGT	65248
Qy	468	GCCATGTTCTGGCGGCTTCTGTGTCTGATCGCTCAGGAGGCGCTCCGCTGCC	527
Db	65247	GCGCTGATGCTGTGGGCTGTGCACCTGTGCGCACCGGCGCGGCGTGGCGG	65188
Qy	528	ACCATGCTCGCTGATCGACCTGCTCAATTTCCGAGTCTGCTCCGCGCGTTCC	587
Db	65187	GTGGCCCGGACCTACCGCGAACTGTCTTCGCGCGGATCAACCGGCTGGTGC	65128
Qy	588	GGCAGGTGGCGGCTGGGTGACTGACACCGTTGCGCATATGTTGCTCTGCTCC	647
Db	65127	GCGACCTCGCTGCTCGGCGGACCTCGCGCGCTGGCGGACGTGGCGCTGGC	65068
Qy	648	GCGCGGAGACTTCTCTGCGACCGGACCGGCGGCTTGAACGCGCGAGAGGCT	707
Db	65067	GCGAGGCGACCGTGGCGGATGTCGCGCGGAACTTCGCGCGCGCGCGCTGCC	65008
Qy	708	CGCGAGCGGTGCAACCGCTCGATCTCTCCCATCGCGATGACCTGGCGCTGGT	767
Db	65007	GAGGCGCGCATCGAGCGGCTCGAGTGGCCGAGAAGGAGGCTTGGCGCTGAT	64948
Qy	768	ACCTCCGCATGACCGGATCGCGTGTGTGATGCTCAGCGCTGCCCATCTCGG	827
Db	64947	ACCGACGCGATGCTCGGCGATGCTGTGTGCTGCCACGATCTCGCGCGTCT	64888
Qy	828	TGGCGGTGGGTTGACGCGCTTCTTCTCGGAATGTCTGAGAGCGCGACGAGG	887
Db	64887	CTGGCGGAGTGAACCGCGCGATGAGCTGGAAGCGCTGTGGGACCGCAAGGT	64828
Qy	888	GCCCGGCGACTGTCCGACTCGGCGCGCATCCCGGACAGAGGACCGCGAGG	947
Db	64827	GCCCGGCGACTGCGAGCGCTGGCGCGGACCCCGGCGGCGGTGGCGGCTG	64768
Qy	948	GCGCGCGCGTGAGCGGCGCGGCGGTGTTCGCGGACGCTATTTGCGACGCG	1007
Db	64767	TCCCGGCTGTCTCGGCGACTCACCCATCTGTGGCGACGCCACCGCGGCT	64713
Qy	1008	GACGCGCGCGATATCGGACGCGGCGGAGCGGCGGAGGATGCTACAGCTCG	1067
Db	64712	-----CCGTGGTGCAGGAGCGCTACTCTCGCTCGCTGC	64681

Qy	1068	GCTCGCAGGTTCTCGGGCGGGCTTCGACACGCTCCGATCGCATCGGGTGTCTG	1127
Db	64680	GCGCGCAGGTCGCGCGCGCGCGCAGGAGACACCTTCGCGCAACCGGTTCG	64621
Qy	1128	ATCAGAGTGAACCGGTTGACCGCAATCCGCTGTTTCGCGCCGATGGCAGCGT	1187
Db	64620	TGGGAGCTGGCGCGCGCGTGGACCAACCCGCTGTGACGCTGGACGGCGGT	64561
Qy	1188	CTGCACGGGGCAATTTTCATGCGGCGAGATGTGGCGGTGACGTCGCGATCG	1247
Db	64560	AAC-----GGCAACTTCCACGCGCGCGCGTGGCTACGCTGCTGGACTTC	64507
Qy	1248	GCCCTCACCCTTCGCGGGGCTTTCGCGAGCGCAGATTCACGCTCTGACAGT	1307
Db	64506	GTGTGCGCGAGCTGGCGAGCATCAGCGAACCGCCACCGATCGTTCTCTGAC	64447
Qy	1308	CTGAACCGTGGGCTGCGCCCTTCTCTCACCGGGCGCGCGGTTGAAATTCGG	1367
Db	64446	GCGAGCACGCGCTTCTGCGCGCTTCTTGGCGACGACCCCGCGG---TGAC	64390
Qy	1368	ATGGCGCAGAGTGAACCGCGACCGCGCTCTTGGCGGAG---ATGCGAGCAC	1424
Db	64389	ATGATCGCCAGTACACCCAGCGCGCATCTCTCCGAGCTCAAGCGGCTGCG	64330
Qy	1425	GCCTCGATCCATTCGATCTCCAGAACCGCGCCCATCAGGATGTGCTCGCT	1484
Db	64329	GCCAGCGTGAATCCATCTCCGCTCTCTCGCGATGACGAGGATCAGCTCT	64270
Qy	1485	ATCGCCGCGCGCTCTCTCGCGAGAAATCAGCCGTTGGGCGGAGATCTTTG	1544
Db	64269	TCGGCGCGCGCAAGTCTGCGCGGCGCTTCGACGCGCTCACCGCGCTGCG	64210
Qy	1545	GCTCTCTCTTTGCAAGCTGCGAGCTGCGC 1577	
Db	64209	GCGCTGACCGCGCGCGCTCTGCGC 64177	

RESULT 12	STMHUTH	2946 bp	DNA	linear	BCT 26-APR-1993
LOCUS	Streptomyces griseus histidine ammonia-lyase (huth) gene, complete cds; ORF, partial cds.				
DEFINITION	M77841.1	GI:153316			
ACCESSION	huth; histidine ammonia-lyase; histidine ammonia-lyase				
VERSION	nonoxidative deamination.				
KEYWORDS	Streptomyces griseus				
SOURCE	Streptomyces griseus				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
REFERENCE	1 (bases 1 to 2946)				
AUTHORS	Wu, P. C., Kroening, T. A., White, P. J. and Kendrick, K. E.				
TITLE	Purification of histidine from Streptomyces griseus and nucleotide sequence of the huth structural gene				
JOURNAL	J. Bacteriol. 174 (5), 1647-1655 (1992)				
PUBMED	1537807				
COMMENT	Original source text: Streptomyces griseus (library: lambda EMBL4; NRRL-B-2682) DNA.				
FEATURES	Location/Qualifiers				
source	1..2946				
	/organism="Streptomyces griseus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1911"				
	/tissue_lib="lambda EMBL4; NRRL-B-2682"				
	<1..727				
	/note="ORF upstream of huth"				
	/codon_start=2				
	/transl_table=11				
	/protein_id="AAA26768.1"				
	/db_xref="GI:153317"				
	/translation="IVLHRAWGELYVARQAGQVFDRAADATVLAUVASGIAOT				
	ERLVEVRKLAFTDPLTGLANRAVDVRLDEAMERHVDATVSVSLVCDLNLKAVNDT				

CDS

HGHVAGDRLIERFSGVLSLGCAMLPEALAAARLGGDEFCLLTAGPPADAVVGVATELCD
RAAVILGDCVAGCAVSTGDPVIGPVRSAARLRLADAAQYRAKAAAPGPVWAGRGE
VLELADSPKPSAHDHRRRLGNRP"

gene

835..2385

CDS

/gene="hutH"

835..2385

/gene="hutH"

/codon_start=1

/evidence=experimental

/transl_table=11

/product="histidine ammonia-lyase"

/protein_id="AAA26769.1"

/db_xref="GI:153318"

/translations="MDMHTVVVGTSTTAEDVVAVARHGARVELSAAVEALAAARLI
VDALAAKPEPVGVSTGFGALASRHIGTELRAOLQNIIVRSHAAAGMPVREVVRAL
MEFLRTVASGHTGVPEVQNTWADVLNAGITPVVHEYSLGSGDLAFLSHCALTIM
GEGEAGPGDVTAPAGELIAAGHAPVEUREKEGLALLNGTDGMLGMVWALADHLRL
YTSADITAALESALSLGTDKVLAPELHAIRPHFGQGVADNMSRVLAGSLGTGHHQDD
APRVQDAYSVRCAPOVNGAGRTDLHAALVAGRELASSVDNPPVLPDGRVESNGHPG
APVAVYDLFLAIVAADLGSICERRTDRLLDKNRSHGLPPFLADDDAGVDSGLMIAOYTG
AALVSMKRLAVPASADSIPSSAMQEDHVMGWSAAKRLRTAVDNLARIVAVELVYAT
RALELRAREGLTTPAPASEAVVAALRAAGEGPDRFLAPDLAAADTFVREGLVAAV
EPTVGTPLA"

ORIGIN

Query Match 12.1%; Score 223.8; DB 1; Length 2946;
Best Local Similarity 49.0%; Pred. No. 1.1e-19; Indels 48; Gaps 4;
Matches 754; Conservative 0; Mismatches 737;

Qy	213	GACAGGCCCATGCGTGGGAGCGGCGCGCGGATGTCCTTGGCCCTCCGGGCGC	272
Db	880	GAGAGCTGTCGCGGTGGCCGCGCCACGCGCGCGGTTCGAGCTCTCCGCGCGCGGTG	939
Qy	273	GACCGGTGCGGTTCGGAAGCGCGCTCGGCGCTGTATCCGCGAGGCGCGCATGTC	332
Db	940	GAAGCCCTGCGCGCGCGCTCTCATGTGAGCGCCCTCGCCGCCAACCCGAGCGGTC	999
Qy	333	TACGACTGACAAACCGGCTTCGTCCTTTCGGAACCGCTGATCTCAGGTGAGAATGTC	392
Db	1000	TACGGTGTCTCCACCGGCTTCGGCGCTTCGCGCCCTCGCGCCGCCACATCGGACGGA	1059
Qy	393	CGAACGCTGACGGCCAACTTTGTCATCATCTGSCACAGCGCGTGGGACCGGTGTC	452
Db	1060	GCGAGCTCCAGCGCAACATGTCGCTCGACGCGCGCGCATGGGCCCGCGGTGAG	1119
Qy	453	TGAGACGCGCGCGCATGTTCTGGCGGTGTGTGTGTCATCGCTCAGGAGGACCTCC	512
Db	1120	CGGAGGTCTGTCGCGCGCTGATTTCTCCGGCTGAAGACGGTTCGCTCGGCGCACAC	1179
Qy	513	GGTGCCAGCGGGACCATGCTCGCTCGATCGACCTGCTCAATTCGAGCTGCTCCG	572
Db	1180	GCGGTACGCCCCGAGGTGCGCGACACATGGCCGACGTGTGAACGCGGGGCAACAC	1239
Qy	573	GCGCTTCCAGCGCGCGCGCGCTCGGGTGCACCTGCACCGCTTGGGCAATG	632
Db	1240	GTGCTCACGAAATAGGCTCGCTCGGTGTCTCGGCGACCTCGCCCGCTCTCGACTG	1299
Qy	633	GTGCTTGTCTCCAGCGCGCGGAGACTTCTCTGAGCCGCGGACGCGGCTTGACGGC	692
Db	1300	GCGCTCACCTCTGATGGCGAGGGCGAAGCGAGGGGGCCGACGCGCACCGTCCGCG	1359
Qy	693	GCAGAGGGCTCCGCGCGGACGCTGCAACCGCTCGATCTCTCCATCGGATGCACTG	752
Db	1360	GGGAAATCTCTCGCGCGCACGGCATCGCCCGGTGAGCTTCGCGAGAGAGGGCGCTG	1419
Qy	753	GCGCTGTCAACGGGACTCTCGCCATGACCGGATCGCGTGTGTGAATGTCTACGCGCTG	812
Db	1420	GCCCTCTCAACGGCAACGAGCGGATGCTCGGATGTGTGTGTCATGGCCCTCGCGG	1479
Qy	813	CGCATCTTCGGCAACTGCGGGGTGGCGTTGACGCGCCCTGTCTTGGGGAATGTCTG	872
Db	1480	AGGAACCTTACACTCGGCGCGACATCACCGCGCGCTGTCTCTTGGAGGGCGCTCTC	1539

Qy	873	CGACACGAGCGATGGGCGCGCGCATGTTCGACCTTCGCGCCCGCATCCCGACAGAGGAC	932
Db	1540	ACGGAACAAGTCTCTGCGCCCGCGAGCTGCACGCGCATCGCCCGCACCCCGGCGTC	1599
Qy	933	GCGGACGAGGCTGCGCGCGCGCGTGGAGCGGCGCGCGGGTGTTCGGGACGTCATT	992
Db	1600	AGCGCGCACAAATGTTCGCGGGTGTTCGCGGTTTCGGGCTGACGGGCTCATCACAGGAC	1659
Qy	993	GCGGACGAGGCTCGACCGCGCGCATATCGGACCGGACCGGAGCGGCGGACAGGATGCC	1052
Db	1660	GACGCCCGCGGTC-----CAGGACGCC	1683
Qy	1053	TACAGCTCGCTGCGCTCCGAGGTTCTCGGGCGGGCTTCGACACGCTCGCATGGCAT	1112
Db	1684	TATTGGTGTGCGTGTGCGCGCCGAGGTCAACGGCGCGCGCGCACACCTTCGACACGCC	1743
Qy	1113	GACCGGTGCTGACGATCGAGCTGAACGCGGTGACCGACATCCGGTGTTCGCGCGCAT	1172
Db	1744	GCCTCTGTCGCGGCGCGCAACTGGCCTCTCTCGTGCACAAATCCCGTGTCTTCCCGGAC	1803
Qy	1173	GCGAGGTGCGCGCTCGACCGGGGCAATTTTCATGGGCGAGCATGTGGCGCTGACGTC	1232
Db	1804	GGAGCGT-----CGAGTCCAAACGGGAATTCACGCGGGCGCGCTCGGCTAGCTCTG	1857
Qy	1233	GATGCGCTCGCCACGCGCGCTCACCGTTCTTGGGGGCTTTGCGGAGCGCCAGATTGCA	1292
Db	1858	GACTTCTCGGATCGTTCGCGCGCGACCTCGGCTCGATCTGCGAGCGCGCACCGACCGG	1917
Qy	1293	CTGACAGATGAAGGCTGAACCGTGGGTGCGCCCTTCTCTCCACCGGGGCCCGCGGG	1352
Db	1918	CTGCTCGACAAGAACCGCTCGCACGCGCTGCGCCGCTTCCT---CGCGGACGACGCGGG	1974
Qy	1353	TTGAATTCGCGCTTCATGGCGCACAGGTGACGGCGACCGCGCTCTTGGCGCGAGAT---	1409
Db	1975	GTGCACTCGGCGCTGATGATGCGCCAGTACACCCAGCGCGCTTGTGTCAGCGAGTGAAG	2034
Qy	1410	CGAGCACGCGGACTGCTCGATTCATTCGATCTCCACGAAACCGCCCAATACGAGTGTG	1469
Db	2035	CGGCTCGGCTCCCGCTCCGCGACTCATCCGCTCTCCGCGATGACGAGGAGCCAC	2094
Qy	1470	GTCTCGTTGGGACCATCGCGCGCGCTCTGCGCGGAGAGATCGACCTTGGGCGGAG	1529
Db	2095	GTCTCCATGGATGTGTCGCGCGCGCGCAAACTCCGTACGCGCGTGGACAACTGGCGCGG	2154
Qy	1530	ATCCTTGGCATCTCTGCTCTTTCGCAACAAGTGGGAGCTGCGCTCGCGCAGCGGC	1589
Db	2155	ATCGTCGCGTGAAGTGTACGCGCGGACCCGCGCCATCGAGCTTCGCGCGCGGAGGC	2214
Qy	1590	CTAGACGGGTGTCTCCCGCGGGAAGAGCTGGTGCAGGCCCTTCGCGGAGCAGATTCCCG	1649
Db	2215	CTCACCCGCGCGCTCGGAGGCGGTGTCGCGCGCTGCGCGCGCGCGCGCGCGAG	2274
Qy	1650	CCGCTTGAGACGCAACCGGCGCTTGGGACAGGAAATTCGCGCGCTTGTGTCGACCTTTG	1709
Db	2275	GGCCCGGCGCGGACCGCTTCTTCGCGCGCGGACCTGCGCGCGCGCGCGGACGCTG	2334
Qy	1710	CAGCAATCTCCCTGAGCGCGCGCTCAGGTCTGTGCG	1748
Db	2335	GAGGGCGCTGTGTCGCGCGCTGAGGCGCGCTCAGCGGG	2373

RESULT 13

BA000040_68/c

WPCOMMENT

Sequence split into 91 fragments			LOCUS	BA000040	Accession	BA000040
Fragment Name	Begin	End				
BA000040_00	1	110000				
BA000040_01	100001	210000				
BA000040_02	200001	310000				
BA000040_03	300001	410000				
BA000040_04	400001	510000				
BA000040_05	500001	610000				
BA000040_06	600001	710000				

Qy	1002	AGGCTCGACCGCGCGGATATCGGACCGGAGCCGGAGCGGGGAGGATGCTTACAGCGTG	1061
Db	65112	-----CGAAACCGTGCAGGATCCTATTGCGTG	65085
Qy	1062	CGCTGCGCTCGGAGGTTCTCGGGCGGGCTTCGACACGCTCGCATGGCATGACCGGTG	1121
Db	65084	CGCTGCGAGCGCGAGGTCGCGAGGGCGCGCTCGATCTGATCAACGAGATCGCGGAGCC	65025
Qy	1122	CTGACGATCGAGCTGAACCGGCTGACCGCAATCCGGTGTTTCGCGCCGATGGCAGCGTG	1181
Db	65024	CTGATCGTCAGGCCAATGCGCTCACCGACAACCGCTCGCTCTGCTCGAGACGGCGAG	64965
Qy	1182	CCGCGCTCGACCGGGGCAATTTCAATGGGCCAGCATGTGCGCTGACGTCGATGCGCTC	1241
Db	64964	ATCGTCTC---CGCGCGCAATTTCCACCGCGAGCCGCTGGCGTTTGGCGCGCATCGATC	64908
Qy	1242	GCACGCGCGTACCGTCTTGCGGGCGCTTGCGGAGCGCCAGATTGACATCTGACAGAT	1301
Db	64907	GCGCTCGCGCTGTCGGAGATCGCGCGCATCAGCGAGCGCGGCATCGCGACGCTGGTCGAT	64848
Qy	1302	GAAGGCTGAACCGTGGGCTGCCCGCTTCTCCACCGGGCGCCCGCGGTTGAATTCC	1361
Db	64847	CCGCGCTGAATTTTCGCTCTGCGCGCGTTCTCAOGCCGATCCCGCG---TCAATTCC	64791
Qy	1362	GGCTTCATGGCGCACAGGTGACGGGACCGCGCTCTCGCCGAG---ATGCGAGCCACG	1418
Db	64790	GGCTTCATGATCGCGAAGTGACGGCGCGCGCTCTATCCGAGAACACGACGCGTGCG	64731
Qy	1419	GGACCTGCTGATCCATTGATCTCACGAAACCGCGCCATCAGGATGCTGCTCGCTT	1478
Db	64730	GCTGCTGCTCGATTCGATTCGACGCCACCGCGCCCAATCAGGAAGACCATGTGTGCGATG	64671
Qy	1479	GGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCGTTGGCGGAGATCCTTGGG	1538
Db	64670	GCGCGCATGCGGCGCGCGCTCGCCGATATGCGGAGCAATCTCGCGCCATCTCGCG	64611
Qy	1539	ATCCTGCTCTCTGTTTGCACAGCTGCGGAGTGGCG	1577
Db	64610	ATCGAGCTTCTGTCGCGCGCAAGCATTCAGCTGGCG	64572
RESULT 14			
AF319998			
LOCUS	49736 bp	DNA	linear BCT 28-MAY-2001
DEFINITION	Stigmatella aurantiaca myxalamid biosynthetic gene cluster,		
complete sequence.			
ACCESSION	AF319998		
VERSION	AF319998.1		
KEYWORDS	GI:14210834		
SOURCE	Stigmatella aurantiaca		
ORGANISM	Stigmatella aurantiaca		
REFERENCE	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;		
AUTHORS	Cystobacterineae; Cystobacteraceae; Stigmatella.		
TITLE	1 (bases 1 to 49736)		
JOURNAL	Silakowski, B., Nordsiek, G., Kunze, B., Blocker, H. and Muller, R.		
PUBMED	Novel features in a combined polyketide synthase/non-ribosomal		
REFERENCE	peptide synthetase: the myxalamid biosynthetic gene cluster of the		
AUTHORS	myxobacterium Stigmatella aurantiaca Sgal5		
TITLE	Chem. Biol. 8 (1), 59-69 (2001)		
JOURNAL	1182319		
PUBMED	2 (bases 1 to 49736)		
REFERENCE	Silakowski, B., Nordsiek, G., Bloecker, H. and Mueller, R.		
AUTHORS	Direct Submission		
TITLE	Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig		
JOURNAL	38124, Germany		
FEATURES	Location/Qualifiers		
source	1. .49736		
	/organism="Stigmatella aurantiaca"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:41"		
	complement(2..1306)		
	/notes="ORF1"		
	/codon_start=1		
CDS			

	/transl_table=11	
	/product="putative aspartokinase"	
	/protein_id="AAK57182.1"	
	/db_xref="GI:14210835"	
	/translation="MEPIGSAQAGRGTSALALIVOKYGGTSGVDTERRLNKAKRCIA	
	AQAGHDVVVVSAMSGETNRLLKLVSQITERNEREQDVVATGQVSGVSLVALAIO	
	TQRRKAVSLGHQVRIITDTSFKARIKSIDAERIIVDALKKNIIVVAGFGQDGEHN	
	VTLRGAGSDTTVALAALAKADACEIYTDVGVYTTDPNMVPAAKSERIAYEEMLE	
	LAAGVAKVQIRSVFAMKYKPLWVKSSITDDPDGPTLVCEADAEMDLVSYADKN	
	EASIALRGVDPVGAAKIFGALDEQSIIVDDLIIVQNSKDGRTDVTFTVGKDLTKAK	
	DVVKTKAVSKAEVETDDQVKSVIGVGRNHSVAAKMTFTVLAGAGVNVQMIST	
	EIKVSCVHSNTELAVRSLHTAFGLDKPEAGAVSEVSLKG"	
	1363. .2895	
	/notes="ORF2"	
	/codon_start=1	
	/transl_table=11	
	/product="putative histidine ammonium lyase"	
	/protein_id="AAK57183.1"	
	/db_xref="GI:14210836"	
	/translation="MSRPRLNDGDTLKEELQVARTVTVELAPAAARVKAARDL	
	VDRVAAGDTPSYGINTGFTGLAEVRIDKKDLRELQNLILSHAAAGVSGPLPEARVL	
	LILRCNVLAGYSGIRPETLALALEMLNRDVPVPERGVSAGDGLAPLHALLVFI	
	GEGEAFYKGERLPAQALERAGLKPVLEAKGLALVNGTOAMCAVGTLLQRAEMLA	
	DIADLAGMTLGLGSHKPFIPETODVRAHQKQCAAHRELRLDADLSALVEHNGS	
	KVQDPYSLRCMPQVHGAREGLSFARRILEVEINSATDNPVFEVETRIIVSGNFHQ	
	PVSLADVAAMALQLSAISERRVEQLVNPALSGLPFLAKNSGLNSGFMIQVTSAA	
	LVAESVLGHPASVDSIPSSAGREDHVSMTAALKGRQVADFTRSCLATLELVAQA	
	LDYQRPTRAGKGPQAAYELIRSKIPTMEKXDRHLRDIAAVALIDSGELINAVRTATR	
	GQ"	
	complement(3275.. .7822)	
	/genes="mxA"	
	complement(3275.. .7822)	
	/genes="mxA"	
	/notes="non-ribosomal peptide synthetase; contains a	
	condensing, adenylation, and PCP reductase domains"	
	/codon_start=1	
	/transl_table=11	
	/product="MxA"	
	/protein_id="AAK57184.1"	
	/db_xref="GI:14210837"	
	/translation="MSVRELLTTLAAKGVRLLEGDKLOVEASKAVLTPELRAELSAN	
	KADILAFLRKHSRETSBELGQPLARPEVILPSSQERLWFLDRSPGTSQYNVHLG	
	LRAGALDITKALRRSDILVRHEVLRTHFFVEGSPQVIVSPDTGVHAIIVELPGL	
	TSQREETIYRRSDLAQQPFDLAGKPLLRVTVALGPDPDFALFVTKHHIITDWSLG	
	VFLQELMYASTVFRGQAALPVSMQFADSLRERAWLAGESGASERAYWKGKGL	
	PIQLPVDHAVSTSHRGATVPLALSALRELASRGCSFTVLPAATFALIH	
	RYSGQVDFGVTVIANRGSVPADLIGFIANTLALCDLSEPTFSCWLAHAKVVLBA	
	LDYQAPFSEVQTVGASRDGGLNPLVRACFTLESIPAPTLIDLPGTWSFSLAPGDS	
	VEGVAKFELSILAASEKGLAGMLEYSREVDASTVERMVGHFQVILLESIVAHPEVPL	
	SKLPLLTABERGLLNDWNGPVLDDVPALCMHELVAQAERTPOAVVVSQKTLTVAE	
	LNRRANQLAHLRLRIGIQKEERVGLCVERTEDIVIGLAILKAGGAVPLDPAYPKER	
	LALILEDQVPLLTQQLVPELPATQARVVCIDKDWPTIGAESDTPNERTIAPETA	
	YLITYSGTGKPKGWIEMHNVAFLIWMSPKELASTSICPDLSPELIFT	
	PLCCGAKVIAKNALSELPELPAAREVILINTVPSAMGALLRSCGVSSVALVNLAGEA	
	LAGALVDVQLDHRVDNVLNPGSETTYSTFTRVNRGQPTTIGRPVGNQTYALDS	
	NREPMIPGVEVYIGMGVARGYGRPELTAEFRVSPFGGSEARLYRTGLARWL	
	PDQLSVLRGMDHOVKRFRIBLGEIGAVLMEHSGIRDAVVRVVRGLAGDKQLVAV	
	VEGEKAPAEIRLDYLKSLPEYMPVFLFVGLDALPTNGKVDRAALPAPERHSG	
	PAKEHAVPTPGEEELAAIWRQVIGVEQIGAHDNFEELGHGSLLYRVVLVLSASCA	
	DIPRLALQATLEENARVAEAKTGLSPAHDTVMEADADLDAIEALGKALPVTVG	
	ALRTILLTATGFLGAPLLEELCRRTDARIYCLVRSKTQEGMNRIRKNLESYLMNE	
	ALAPRIYVPRGDTQGLLSEKFORLSEEDAIYHNGALVNFPLYPESMRANVLG	
	TRILRLATRTRIKPLHYVTSVSLPLGRKAPIREDEPLEGSSLVGGVAGSKWAEK	
	LYPEASRGGLPVTILRPGVTHGSRGTGAWNTDLCVRLTKGCVRMVAPSVADLLDT	
	PDYVSSALVDLSMRPESIGQTHLYNPQVPRADENMYNRAFGYGLRVLPYDQWLS	
	LGSAASSDELGLMLPQQVPPEDRSVGGPRMVVCDSDGTLKALGSGTGTSCPSVDAS	
	ListyLSLVLVHRGLKAEVR"	
	complement(7819.. .11520)	
	/genes="mxA1"	
	complement(7819.. .11520)	
	/genes="mxA1"	
	/notes="PKS; ketosynthase; acyltransferase"	
	/codon_start=1	

1103	Qy	CGCATGGCATGACGGGTGCTGACGATCGAGCTGGAACGCGGTGACCGACAATCCGGTGTT	1162
894	Db	CGGGTTCTCCCGGCGCATCTCTGGAGGTGGAGGTCAACAGCGCGACGCAACCCGCTCGT	953
1163	Qy	TCGCCCGATGGCAGCGGTGCCGCCCTTGACAGGGGGCAATTTTCATGGGCGAGCATGTGGC	1222
954	Db	---GTTTCGGGACACGGAGCGCATCTGTCTGGGCGGCACTTTCACGGCCAGCCCACTCTC	1010
1223	Qy	GCTGACGTCCTCGATGCGCTCGCCACAGGCCGCTCAACGTTTCTGGCGGGCTTTGCGAGGCGCCA	1282
1011	Db	CCTGGCCATGGACGTGTGGCGCATGCGCGCTTGAACGCAACTGTCTGTCCATCAGCGAGCGCGC	1070
1283	Qy	GATTGCACTGCTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTCTCCACCGGGG	1342
1071	Db	CGTGGAGCACTCGTGMAACCCGTCGCTGTCCAACCTGCGGGCGT-----TCCTGGCGAA	1124
1343	Qy	CCCCCGCGGGTTGAATTTCCGGCTTTCATGGCGCGCACAGGTGACGCGGACCGCGCTCTCTGGC	1402
1125	Db	GAACTCCGGGTTGAATCTCCGGCTTCATGATCGCGCAGGTGACACGAGCCGCGCTGTGTGGC	1184
1403	Qy	CGAGATGCGAGCACCGGGA---CCTGGCTCGATGCCATTCGATCTCCACGAAACGCCGCCAA	1459
1185	Db	CGAGTCCCGGCTGCTGAGCCACCCCGCGTCCGTGGATTCGATTTCCGTCATCCCGCGGGCCG	1244
1460	Qy	TCAGGATGTGCTCTCGCTTCGGGACCATCGCGCGCGCTCTCGCGGAGAGAATCGACCG	1519
1245	Db	AGAGGACCACTGTCTCATGGGCGATGACGGCGGCGCTCAAGGGCCGTCAGGTGACGACATT	1304
1520	Qy	TTGGGCGGAGATCCTTTGCGATCCTCGCTCTCTGTCTTGCAAAAGCTGCGAGCTGCGCTG	1579
1305	Db	CGCCCGTTCTGCGCTCGCGATTGAATTCCTTGGTGGCGGCGAGGCCCTTGACATTCCGCT	1364
1580	Qy	CGGCGGGGCTAGACGGGGTGTCTCCGCGGGGAGAGAGCTGGTGAGGCCCTTGCGGA	1639
1365	Db	GCCGCTGAAGCCCGGCAAGGGCGCCCTCGCGGCGTACGAGCTGG-----TGCGCTC	1415
1640	Qy	GCAGTTTCCCGCCGCTTGAGACGGACCGGCCCTTGGGACAGAGAAATTCGCGCGCTTGCTAC	1699
1416	Db	GAAGGTCGCCCAATGGACAAAGACCGGAGCTGCACCGGGAATCGAGCGGTGAGCCA	1475
1700	Qy	GCACCTC 1706	
1476	Db	GCTCGTC 1482	

Search completed: December 10, 2005, 15:10:47
Job time : 9347 secs

THIS PAGE BLANK (USPTO)

Result No.	Query			DB	ID	Description	
	Score	Match	Length				
C	1	309.8	16.7	400	9	B07840	B07840 85367P101C31
	2	122.2	6.6	2151	10	CL972100	CL972100 05IFCC041
	3	119	6.4	2598	4	A103647	A103647 Zea mays
	4	111	6.0	4590	10	CL982033	CL982033 05IFCC046
	5	104.8	5.7	1740	10	CL967485	CL967485 05IFCC040
	6	102.2	5.5	669	9	BX39558	BX39558 EINAG81TF
	7	97	5.2	812	10	CG301209	CG301209 OG3AG33TV
	8	97	5.2	845	10	CG362861	CG362861 CG3A210TH
	9	97	5.2	888	9	CG664163	CG664163 OGNAF84TV
	10	94.4	5.1	737	7	CF842933	CF842933 p8H8023xA
C	11	92	5.0	788	8	DN640249	DN640249 UMC-bend
	12	92	5.0	930	8	CX949440	CX949440 UMC-bcl_0
	13	90.6	4.9	847	8	DR785713	DR785713 ZM BPH000
	14	90.4	4.9	786	6	CB653003	CB653003 OSJNEC03J
	15	90.4	4.9	866	6	CB642693	CB642693 OSJNEB03A
	16	90.2	4.9	856	6	CB654478	CB654478 OSJNEC07A
	17	90	4.9	841	10	CG298372	CG298372 OXGDK64TV
	18	90	4.9	1429	9	B2578760	B2578760 mah2 5990
	19	89.8	4.9	831	9	CG675672	CG675672 OGUES16TV
	20	89.6	4.8	808	6	CB655968	CB655968 OSJNEC09M
21	89.2	4.8	859	6	CB632830	CB632830 OSTIEE11F	
22	88.4	4.8	796	6	CB630090	CB630090 OSTIEE06N	

161:955-962. The library was then ordered around chromosome II of Rhodospirillum rubrum (Choudhary et al., 1994. Journal of Bacteriology, 176:7694-7702). The cosmids were then digested with a variety of restriction enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and DNaseI) and restriction fragments subcloned into the respective multiple cloning site sites of pBluescript SK (-). Note BglII fragments were subcloned into the pBluescript BamHI site. DNaseI fragments were subcloned into the EcoRV site. All subclones were transformed into E. coli XLBlue MRF'. All fragments were then sequenced and the sequences where possible were assembled using the GCG program GELASSEMBLE."

```

ORIGIN
Query Match      16.7%; Score 309.8; DB 9; Length 400;
Best Local Similarity 89.0%; Pred. No. 3.3e-56;
Matches 333; Conservative 0; Mismatches 38; Indels 3; Gaps 1;

QY 545 CGACCTGCTCAATTCGAGCTCGCTCCGGCCGTTCC---AGCCGGCGCACGGTGGGCGC 601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 CGACCTCTCAANNCCGAGATCCCTCCGGCCCGTTNNAANNACGGAAGGTGGGCGC 316

QY 602 GTCGGGTGACCTGACACCGCTTGGGATATGCTGCTCTCAGGGCCGGGAGACTT 661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 GTCGGGTNANCTGACACCGCTTGCANANATGGTCTCTCCCTCAGGNCGGGGAGACTT 256

QY 662 CTTGGACCGGCGGAGCGGCTTGACGGCGCAGAGGGCTCCGGCGGAGCGGTGCA 721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 CTTGGACCGGAGCGGAGCGGCTTGACGGCNCNAGAGGGCTCCGGCGGAGCGTCTCCA 196

QY 722 ACCGCTCGATCTCTCCATCCGATGCACTGGGGCTGTCAAGGGACCTCCGCCCATGAC 781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 ACCGCTCGATCTCTCCATCCGATGCACTGGGGCTGTCAAGGGACCTCCGCCCATGAC 136

QY 782 CGGGATCGGCTGTGATCTCAGCTCCGCGCATCTCGGCACTGGGCGGTGGGCTT 841
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CGGGATCGGCTGTGATCTCAGCTCCGCGCATCTCGGCACTGGGCGGTGGGCTT 76

QY 842 GACGGCCCTCTTGGGAATGCTGAGAGCGCGGACCGAGGATGGCGCGGCACTGTC 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 GAGNCCCTCTCTGCGAATGCTGAGAGCGCGGACCGAGGATGNCNCGGCACTGTC 16

QY 902 CGACCTCGGCGCGC 915
    ||||| |||||
Db 15 CGACCTCGGCGCGC 2

```

RESULT 2
CL972100 2151 bp DNA linear GSS 21-SEP-2004
LOCUS OsIFCC041692 Oryza sativa Expressed Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.

ACCESSION CL972100.1 GI:52398729
VERSION 1
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)

REFERENCE 1 (bases 1 to 2151)
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.

TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559

Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source 1. 2151
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed Library"
/notes="Oryza sativa exon trapped genomic sequences"

```

ORIGIN
Query Match      6.6%; Score 122.2; DB 10; Length 2151;
Best Local Similarity 47.5%; Pred. No. 1e-15;
Matches 660; Conservative 0; Mismatches 678; Indels 51; Gaps 8;

QY 219 GCCCATGCGGTGGCGAGCGCGCGGATGTTCTTGGCCCTCCGGCGCGACCGG 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 GCCGGCGCGCGAGCGCGCGGATGTTCTGAGCTCGACGATCGGCGCGGAGCGC 270

QY 279 TGCCTGCGTCCGAAGCGGGCTCGGCGTGTCTCCGGAGGCGGCCATGTCTACGGA 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 GTCAAGGCCAGCAGCTGTTGCGAACCGCTGATCTCAGGTGAGAAATGTCCGAACG 330

QY 339 CTGACAAACCGGCTTCCGTTGCGAACCGCTGATCTCAGGTGAGAAATGTCCGAACG 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 GTCAACACCGGTTCCGGCCAGCTCGACCGCGGACCAAGAGGGCGCGCTGCAG 390

QY 399 GTGACGGCAATCTTGTCTCATCATCTGG-----CCAGCGCGTGGGACCGGTG 446
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 391 CGGGAGCTCATAGGTTCTCAACGCGCGCGCTTCGGCACCGCGCGCGACGCCACGTC 450

QY 447 CTTGATGAGACGCGCGCGCGCATGTTCTGGCGGCTGTGTGTCGATCGCTCAGGGA 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 451 CTCCTCGGGGTGCCACGCGCGCGGCGATGCTCGTCCGGATCAACACCTGTCTGCAAGGC 510

QY 507 GCCTTCGGTCCACGAGCGGCGACCATCGCTCGCTGATCGACCTGTCTCAATTCGAGCTC 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 511 TACTCCGGCATCCGTTTCGAGATCTCGAGGGCGTCCCAAGCTGTCAACGCCAAGCTC 570

QY 567 GCTTCGGCGCTTCCAGCGCGCGCACGCGTGGGCGCGTGGGTGACCTGACACCGCTTGGC 626
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 AGCCCGTGTCTCGCGCTCCGGGCGACGATCACCGCGTGGGCGGACCTCGTCCCGCTCC 630

QY 627 CATATGCTGTCTGCTCTCAGGCGCGGAGACTTC-----CTGACCGGGAGCGGACG 680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 TACATCGCGGGCTCTGCTCACCGCGCGGAGAACTCCGTGGGTGGCGCCGACGCGAGG 690

QY 681 CGGCTTTCGAGCGCGAGAGGGCTTCGGCGCGGAGCGCTGCAACCGC-----TCGATCTC 734
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 691 AAGGTGAACCGCGCGAGCGGTTCAAGATCGCGGCGATCCAGGGCGGCTTCTTCGAGCTG 750

QY 735 TCCCATCGGATGCACTGGGCGTGTCAAGCGGACCTCCGCCATGACCGGGATCGCGCTG 794
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 CAGCCCAAGAGAGGCTTCGCGATGTTGAAAGCGGCGGCTGGGGCTCCGGCTCCGCTCC 810

QY 795 GTGAATGCTCAGCCTCGCGCATCTCGCAACTCGGGCGGTGGCGTTGACGGCCCTGCTT 854
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 ACGTCTGTTTCGAGGCCAAGCTCTCGCCATCTCGCGAGGTGTCTTCGCGCGGTTC 870

QY 855 GCGGAATGCTCTGAGAGCGCGGACCGAGGATGGGCGCGGCGCACTGTCCGACCTTCGCGCGC 914
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 TCGAGGTGATGACCGGCAAGCGGAGTACACCGACCACTCACCCACAAGCTGAAGCAC 930

QY 915 CATTCGCGAGAGAGAGCGCGGAGGCTTCGCGCGCGCGCTGGAGCGGAGCGCGCGG 974
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 CACCCGGGCGAGATCGAGGCGCGCGCATCATGGAGCACATCTTGAAGGAGCAGCTC---- 986

QY 975 GTGGTCCGCGACGTCAATTCGCGAGCGGAGGCTCGACCGCGGCGATATCGGGAGCGGAGCG 1034
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 987 -----GTACATGAAGGAGCGGAGAGGCT-----CGCGAGCTCGACCCGCTGATG 1032

```


Db 941 TCGAGGTCTATGAACCGCAAGCCGAGTACACGACCACTGACCCACAAGCTGAAGCAC 1000
QY 915 CATCCCGGACAGAGGACCGCGAGCGAGGTGCGCGCCGCGTGGACGCGAGCGCGCG 974
Db 1001 CACCCGGGTTCATCGAGCGCGCGGCATATGAGACATCTCTGGATGGCAGCTC---- 1056
QY 975 GTGTCCCGGCACGTCTATTGCGCGAGCGGAGGTCTGACGCGCGCGCATATCGGGACGAGCCG 1034
Db 1057 -----CTTCATGAACGAGCGCAAGAGTTGAACGAGCTGGACCCGCTGCTG 1102
QY 1035 GAGCGGGGCGAGTGTCTACAGCTGGCTGCGCTCGCAGGTTCGCGGGCGGGCTTC 1094
Db 1103 AAGCCCAAGCAGGACAGTACGCTCCGACGCTCGCGCAGTGGCTGGCGCCCGCCAGATC 1162
QY 1095 GACACGCTCGCATGGCATGCCGGTCTGACGATCGAGTGAACGCGGTGACCGCAAT 1154
Db 1163 GAGTCTATCGCGCGCGCCACCAAGTCAATCGAGCGAGGTCAACTCCGTGAAGCAAC 1222
QY 1155 CCGGTGTTTCGCGCCGATGGCAGCTGCGCGCCCTGACCGGGGCAATTTTCATGGGCGAG 1214
Db 1223 CCGGTCTACGAGCTCCACCGCGGCAAG--GCGCTGACGCGGCACTTCCAGGGCACC 1279
QY 1215 CATGTGGCGTGAAGTCCGATGCGATGCGCCTGCGCACGGCGCTCACCGTTCGCGGGCGCTTGGC 1274
Db 1280 CCCATCGCGGTGTCCATGGAACACCGCGCGCTCGCCATCGCCAACTCGGCAAGCTCATG 1339
QY 1275 GAGCGCAGATTGCACTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCT- 1333
Db 1340 TTGCGCGAGTTCTCCGAGCTCGTCAACGAGTTCTTACAAACACGCGGCTCACTTCAACCTG 1399
QY 1334 --CCACCGGGCGCCCGCGGTTTGAATTCGCGCTTCATGGCGGCACAGGTGAACGCGGACC 1391
Db 1400 GCCGACAGCGCAACCCAGCGCTGAGTACGCGCTTCAAGGGCACCGAGATGCGCATGGCC 1459
QY 1392 GCGCTCTCGCGGAGATGCGAGCCACCGG--ACCTGCGCTGATTCCTANTGATTCACG 1448
Db 1460 TCTTACTGCTCCGAGTCCAGTACTCGGCAACCCCACTCAACCAACGCTGCGAGGCGC 1519
QY 1449 AACCGCCCAATCAGATGATGGTCTGCTGCGGACCATCGCGCGCGCTTCGCGGAG 1508
Db 1520 GAGCAGCACAAACAGGACGTAATCTCCTGGGCGCTGCTGCGGCAAGAACCGCGCGAG 1579
QY 1509 AAGATCAGCGTTGGGCGGAGATCTTTGCGATCTCGCTCTCTGTCTTGACAAAGCTGCG 1568
Db 1580 GCGATCGACATCTGGAAGCTCATGCTCGTCCACCTACATCGTGGCGCTGTGCCAGGCGCTG 1639
QY 1569 GAGCTGCGC 1577
Db 1640 GACCTGCGC 1648

RESULT 4
CL982033
LOCUS OaIFSC046542 Oryza sativa Exprim Library Oryza sativa (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL982033
VERSION CL982033.1 GI:52418552
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 4590)
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
TITLE An analysis of transcriptional regulation of the rice genome and
JOURNAL its comparison to Arabidopsis
COMMENT Unpublished (2004)
Contact: Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
1..4590
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Exprim Library"
/note="Oryza sativa exon trapped genomic sequences"
FEATURES
source
ORIGIN
Query Match 6.0%; Score 111; DB 10; Length 4590;
Best Local Similarity 48.9%; Pred. No. 2.8e-13;
Matches 328; Conservative 0; Mismatches 340; Indels 3; Gaps 1;
QY 267 GCGGCGACCGGTGCGTCCGAGCGGCTCGGCGCTGTCATCCGCGAGCGCGC 326
Db 2635 GCCTACGCGCCTATCCAGCAGAGCGTGGCTCGGTCGAGCAGATCATCGCGAAGACCGC 2694
QY 327 CATGTCTACGGAAGTGAACACCGGCTTCGCTCCGAAACCGCCTGATCTCAGGTGAG 386
Db 2695 ACGGCTACGCGCATCAACCGGTTTCGGCTGCTGCTTCGACCCGATCGCCGTCAC 2754
QY 387 AATGTCCGAAGCTGACGCGCAATCTTTGTCATCATCTGGCCAGCGCGTGGAGACCGGTG 446
Db 2755 GATTGGAAACCTGACGCGTTCAATCTGCTGTACACCGCGCGGCTGGGCGAGCCC 2814
QY 447 CTGACTGACGAGCGCGCGCGCATGTTCTGGCGGTGCTGTCATCGCTCAGGGA 506
Db 2815 ACCGACGACAATCTGGTGGCGCTGATGATGGTGTGAAATCAACACCTGTTCGCGCGC 2874
QY 507 GCCTCGCGTCCAGCAGCGGAGCACCTCGCTCGCTGATCGACCTGCTCAATTCGAGCTC 566
Db 2875 TTCTCGCGCATTCGCTGGAAGTCAATCCAGCGCTGATCGCGTGTCAACCGCAAGTT 2934
QY 567 GCTCGCGCGCTTCCAGCGCGCGGACGCTGGGCGCGTGGGTGACCTGACACCGCTTGGC 626
Db 2935 TATCCGCAATATCCGCTGAAAGGCTCGGTGCGGCGCTTCGCGGACCTGGCGCGCTGGC 2994
QY 627 CATATGCTGCTCTGCTCCAGGCGCGGAGACTTCTTGACCGCGGACGCGCGCTT 686
Db 2995 CACATGAGCTGTGCTGCTGGGCGAAGCGCGGTCTATCAGGCGCAATGCTGCGC 3054
QY 687 GAGCGCGCAGAGGCTTCCGCGCGGACGCGCTGCAACCGCTCGATCTCTCCCATCGCGAT 746
Db 3055 GCGACCGAAGCGTGGCCAGGCGCG--TCTGAACCGCTGACGCTGGCGGCGAAGAG 3111
QY 747 GCACTGCGCTGTGTAACCGGACCTCGCATGACCGGGATCGCGCTGGTGAATGCTTAC 806
Db 3112 GGCCTGGCGCTGCTGAACCGCACCCAGTCTCCGCGCGCTTCGCGTTCGCGCGCTGTC 3171
QY 807 GCCTGCGCGCATCTCGGCAACTGGGCGGTGGCTTGACGCGCGCTTGGCGGATGCTG 866
Db 3172 GACGCGAAGACCTGTTGCGCGCGGACCGTGGCGGCGACGCTGAGGTGGAGCGCGC 3231
QY 867 AGAGGCGGACCGAGGCGATGGGCGGCACTGTCCGACCTGCGCGCGCATCCCGGACAG 926
Db 3232 CTGGGCTCCCGACGCGCTTCGATGCGCGTATCCATGAGGTGCGCGGCGAGCGCGCAG 3291
QY 927 AAGAGCGCGC 937
Db 3292 ATCGAGCGCGC 3302

RESULT 5
CL967485
LOCUS CL967485 1740 bp DNA linear GSS 21-SEP-2004

DEFINITION OSIFCC040011 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL967485.1 GI:52389612
VERSION CL967485.1
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Euphorbiales; Euphorbiaceae; Oryza.
REFERENCE 1 (bases 1 to 1740)
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wang, G., K. S., Deng, X. W. and Wang, J.
TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
1..1740
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences "

Query Match 5.7%; Score 104.8; DB 10; Length 1740;
Best Local Similarity 47.7%; Pred. No. 5.7e-12;
Matches 558; Conservative 0; Mismatches 572; Indels 40; Gaps 7;
427 CCAGCGGCGTGGACCGGTCTTGACTGGACGACGCGCGCGGCCCATGGTTCTGGCGCGTC 486
19 CGGCAACGACGCGGCGCACGTTCTCGCGCGGAGCGCGCGCGCATGCTGCTGCCGA 78
487 TGGTGTGATGCTCA-GGAGCTTCGGTGCCAGCAGGAGGACCATGCTGCTGCTGATC 545
79 TCAACACCTCTCTCCAGGGGTACTCTCGGCGATCCGTTTCGAGATCCTCGAGGCCATCACC 138
546 GACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCCAGCGCGGACCGTGGCGGCTCG 605
139 AAGCTGCTCAATGCAACGTCACGCGCTGCTGCGCTCCGCGCGGACCATCACCGCGCTCC 198
606 GTGACCTGACACCGCTTGGCGCATATGCTGCTGCTCCAGGCGCGGAGACTTCCGTG 665
199 GGTGACTTGGTCCCACTGTCTACATTCGCGGCTCATCACCGCGCGCGAGACTCCGTG 258
666 G-----ACCGGACGCGGCGTGTGACGCGCGCAGAGGGCTCCGCGCGGACGCGTGC 720
259 GCGGTGCGCGCGATGGCGCAAGGTGACCGCGCGGAGGATTCAGATTGCGGCGATT 318
721 AACCGC-----TCGATCTCCCATCGCATGCTGCGGCTGCTGCTCAACGGGACCTCC 773
319 GAGCACGCGCTTCTTCGAGTTGCGAGCTAAGGAAGGTCTTGCCATGTTGACGCGGACGCGC 378
774 GCATGACGCGGATCGCGCTGCTGTAATGCTACGCGCTCGCGCATCTCGGCAACTGGGCG 833
379 GTCGCTCTGGCTTGGATGACCGCTGCTTGTGAGGCGCAAGTCTTGGCATCTCTGCC 438
834 GTGGCTTACGCGCTCTCTGCGGAATGTCTGAGAGCGCGGACCGAGGATGGGCGCG 893
439 GAGGTCTGTGCGCGTGTCTGTGAGGTTCATGACCGCGCAAGCGGAGTACACCGACAC 498
894 GCATGTCTGCGGCTGCGGCGATCCCGACAGAGGAGCGCGGAGGCTGGCGCG 953

499 CTGACACACAAGCTGAAGCACCCCTGGACAGATCGAGGCTGCGCCATCATCATGAGCAC 558
954 CGCTGGACGCGAGCGCGGGTGGTCCGGCACGCTCATTCGCGAGGAGGCTGACGCG 1013
559 ATCTTGGAGGCGAGCTCGTACATG-----AAGCTGGCGAAGAGCTC 600
1014 GCGATATCGGAGCGAGCGCGGAGCGGCGGAGGATGCTACAGCTGCGCTGCGCTCG 1073
601 GCGAGCTCGACCGCTTATGAGCGGAGGAGGAGGATGCGCTCGGACGCTGCGCG 660
1074 CAGGTTCTCGGCGGCGGCTTCGACACGCTCGCATGCGCATGCGGATGCTGACGATCGAG 1133
661 CAGTGGCTCGGCGGCGGCGGAGGATGCGCTCGCTTCGCGCACCAAGCTCGATCGAGCGGAG 720
1134 CTGAACGCGGCTGACGCAATCCGGTGTTCGCGCGGAGGAGGCTGCGCGCTCGCTGAC 1193
721 ATCAACTCCGTCAACGACCAACCCGCTGATGCTGCTCCGCTGGCAAG---CGGCTTCAC 777
1194 GGGGCAATTTTCATGGGCGGAGCGGCTGCGCTGCGCTGCGCTGCGCTGCGCACGCGCTC 1253
778 GGTGGCAATTTTCAGGCGGCGGCGGAGGATGCGCTGCGCTGCGCTGCGCTGCGCTC 837
1254 ACCGTTCTGGGCGGCGCTTTCGCGGAGCGGCGGAGGATGCGCTGCGCTGCGCTGCGCTG 1313
838 GCTGCCATCGCAAGCTCATGTTCCGCGAGTCTCTGAGCTCGTGAACGACTTCTACAC 897
1314 CGTGGCTGCGCGGCTTCTCTCCACCGCGGCGCGG---CGGGTTGAATTCGCGCTTCATG 1370
898 AACGGCTTCTCTCCAACTGTCCGCTGCGGAGCGCAACCCGAGCTTATGGTTCAAG 957
1371 GGCGCACAGGTGAGCGGCGGCGGCTGCTGCGGAGGATGCGGAGCGGCGG---ACTGCC 1427
958 GCGCGGAGATCGGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
1428 TCGATCCATTCGATCTCCACGAACCGCGCGGAGGATGCTGCTGCTGCTGCTGCTGCTGCT 1487
1018 ACCAACCATGTCAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1077
1488 GCGCGCGCTGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1547
1078 TCCTCCAGGAAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137
1548 CTCTGCTTTCGACAGCTGCGGAGCTGCGC 1577
1138 ATCGCGCTGTGCCAGGCGCATCGACCTGCGC 1167

RESULT 6
BZ396558
LOCUS EINAG81TF EI_10.12_KB Entamoeba invadens genomic clone EINAG81,
DEFINITION genomic survey sequence.
ACCESSION BZ396558
VERSION BZ396558.1 GI:30243105
KEYWORDS GSS.
SOURCE Entamoeba invadens
ORGANISM Entamoeba invadens
REFERENCE 1 (bases 1 to 669)
AUTHORS Wang, Z., Samuelson, J., Clark, C. G., Eichinger, D., Paul, J., van Dellen, K., Hall, N., Anderson, J., and Loftus, B.
TITLE Gene discovery in the Entamoeba invadens genome
JOURNAL Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
PUBMED 12798503
COMMENT Other GSSs: EINAG81TF
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
DNA was provided by Daniel Eichinger


```
genomic survey sequence.
CG362861.1 GI:34280128
GSS.
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1. (bases 1 to 845)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reanick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: G3A210TV
Contact: Cathy Whitelaw
TIGR
7712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: methylation filtered.
Location/Qualifiers
1..845
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0766A20"
/clone_lib="ZM 0.7-1.5_KB"
/note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 5.2%; Score 97; DB 10; Length 845;
Best Local Similarity 51.2%; Pred. No. 2.6e-10;
Matches 285; Conservative 0; Mismatches 260; Indels 12; Gaps 2;

Qy 427 CCAGCGCGTGGGACCGGTGCTTCTGACTGGACGCGCGCGCCATGGTTCGTGGCGGTC 486
Db 620 CCGGGCGGACGGCCACGTGTGCGGCCGAGGCGACGCGCGCGGCGATGTCGTGGCA 561
Qy 487 TGGTGTGATCGCTCAGGAGCGCTCCGGTGCACGAGGAGGACCATCGCTCGCTGATCG 546
Db 560 TCAACACCCCTCCTCAGGGCTACTCCGGCATCCGCTTCGAGATTCTGGAGGCCATCGCA 501
Qy 547 ACCTGCTCAATTCGAGCTCGCTCGGGCGGTTCCAGCGCGGCGACGGTGGGCGGTCGG 606
Db 500 AGCTGCTCAAGCCAAAGTCACGCGGCTGCTGCGGCTCCGGGCGACCATCACCGGTCGG 441
Qy 607 GTGACCTGACACCGCTTTCGATATGTTGCTCTGCTCCAGGCGCGGGAGACTTC---- 662
Db 440 GCGACCTCGTCCGCTCTCTACATGCCGCGCTCATCAGGGCGCGCAGAACTCCGTGCG 381
Qy 663 --CTGGACCGGGACGGGACGGCTTGAACGGCGCAGAAAGGCTCCGGCGCGGACGGTGC 720
Db 380 CGGTGGGCGCCGACGGCAGGAAGGTGGGCGCGCGGAGGCTTCAGGATCGCGGCATCG 321
Qy 721 AACCG-----CTGATCTCTCCATCGCATGCTGCGCGCTGCTCAACGGGACCTCGG 774
Db 320 AGCAGGGGTTCTTCGAGTGCAGCCCCAAGGAGGGGCTCGCCATGTTCAACGGGACCGCG 261
Qy 775 CCATGACCGGATCGCGTGTGTAATGTCACGCTGCGCGCATCTCGGCAACTGGGCGG 834
Db 260 TGGGCTCGGGCTCGCTCCACCTGCTCTTCGAGGCCAAGTTCCTGCGCTATGGCGG 201
Qy 835 TGGCGTTGACGGCCCTGTTGCGGAATGTTCTGAGAGGCGGACGAGGCATGGGCGCGG 894
Db 200 AGGTCAATCGCGGTGTTCTCGCAGGTTCATGACCGGCAAGCCGAGTTTCAACCGACACC 141
Qy 895 CACTGTCCGACCTGGGCGGCATCCCGGACAGAAAGGACGCCGCGAGCGGTGCGCGGCC 954

Db 140 TGACGCACAAGCTGAAGCACCAACCCCGACAGATCGAGGCGCGCTATCATGAGCAGC 81.
Qy 955 CGGTGGACCGCAGCGCG 971
Db 80 TCCTGGAAGGCAGCTCG 64

RESULT 9
CC664163
LOCUS
DEFINITION
OGNAP84TV ZM.0.7-1.5 KB DNA linear GSS 19-JUN-2003
genomic survey sequence.
CC664163
VERSION
CC664163.1 GI:32068134
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1. (bases 1 to 888)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reanick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGNAP84TH
Contact: Cathy Whitelaw
TIGR
7712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: methylation filtered.
Location/Qualifiers
1..888
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0325M24"
/clone_lib="ZM 0.7-1.5_KB"
/note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 5.2%; Score 97; DB 9; Length 888;
Best Local Similarity 51.2%; Pred. No. 2.6e-10;
Matches 285; Conservative 0; Mismatches 260; Indels 12; Gaps 2;

Qy 427 CCAGCGCGTGGGACCGGTGCTTCTGACTGGACGCGCGCGCCATGGTTCGTGGCGGTC 486
Db 304 CCGGGCGGACGGCCACGTGTGCGGCCGAGGCGACGCGCGCGGCGATGTCGTGGCA 363
Qy 487 TGGTGTGATCGCTCAGGAGCGCTCCGGTGCACGAGGAGGACCATCGCTCGCTGATCG 546
Db 364 TCAACACCCCTCCTCAGGGCTACTCCGGCATCCGCTTCGAGATTCTGGAGGCCATCGCA 423
Qy 547 ACCTGCTCAATTCGAGCTCGCTCGGGCGGTTCCAGCGCGGCGACGGTGGGCGGTCGG 606
Db 424 AGCTGCTCAACGGCAACGTCACGCGGCTGCTGCGGCTCCGGGCGACCATCACCGGTCGG 483
Qy 607 GTGACCTGACACCGCTTTCGATATGTTGCTCTGCTCCAGGCGCGGGAGACTTC---- 662
Db 484 GCGACCTCGTCCGCTCTCTCTAATCCGCGCTCATCAGGGCGCGCAGAACTCCGTGCG 543
Qy 663 --CTGGACCGGGACGGGACGGCTTTCACGCGCGCAGAAAGGCTCCGGCGCGGACGGTGC 720
Db 544 CGGTGGGCGCCGACGGCAGGAAGGTGGGCGCGCGGAGGCTTCAGGATCGCGGCATCG 603
Qy 721 AACCG-----CTGATCTCTCCATCGCATGCTGCGCGCTGCTCAACGGGACCTCGG 774
```

```

Db      604 AGCAGCGGTCTTCTGAGCTGCAGCCCAAGAGGAGGCGCTCGCCATGGTCAACGGCACCGCCG 663
Qy      775 CCATGACCGGATCGCGCTGGTGAATGCTACAGCGCTCGCGCATCTCGGCAACTGGGCGG 834
Db      664 TGGGCTCGGCTCGCTCCACCGTCTCTTCGAGGCAAGTTCCTCGCGTCAATGCGCG 723
Qy      835 TGGCGTTGACGGCCCTCTTTCGGAATGTTCTGAGAGCGCGGACCGAGGCATGGGCCCGG 894
Db      724 AGGTCACTCTCGCGGTGTTCTGCGAGTCTATGACCGCAAGCCGAGTTCACCGACCA 783
Qy      895 CACTGTCCGACCTCGGCGCGCATCCCGGACAGAAGGACGGCGAGCGTGGCGCCCC 954
Db      784 TGACGCAAGCTGAAGCACCAACCCCGGACAGATCGAGGCGCGCGCTATCATGAGCAG 843
Qy      955 GCGTGGACGGCAGCGCG 971
Db      844 TCCTGGAGGCGAGCTCG 860

RESULT 10
CF842933
LOCUS   CF842933                737 bp      mRNA      linear      EST 30-OCT-2003
DEFINITION pSHB023x13f USDA-IFAPs:Expression of Phytophthora sojae genes
           during infection and propagation_sHB Phytophthora sojae cDNA clone
           sHB023A13 5, mRNA sequence.
ACCESSION CF842933
VERSION   CF842933.1 GI:38058587
KEYWORDS  EST.
SOURCE    Phytophthora sojae
ORGANISM  Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
           Phytophthora.
REFERENCE 1 (bases 1 to 737)
AUTHORS   Tyler,B. Not Published
TITLE     Unpublished (2003)
JOURNAL   Contact: Tyler B
COMMENT   Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel.: 540-231-7318
Email: bmt Tyler@vt.edu
PCR Primers
FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 023 row: A column: 13
Seq primer: BK reverse primer
High quality sequence stop: 737.
Location/Qualifiers
FEATURES
source
1..737
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sHB023A13"
/tissue_type="mycelium"
/cell_line="p6497"
/dev_stage="48 hr. post infection stage"
/lab_host="Soybean plant"
/clone_lib="USDA-IFAPs:Expression of Phytophthora sojae
genes during infection and propagation_sHB"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN
Query Match 5.1%; Score 94.4; DB 7; Length 737;
Best Local Similarity 50.0%; Pred. No. 9.3e-10;
Matches 262; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

Qy      310 TCATCCCGGAGGCGCGCATGCTACGGACTGACAAACCGGCTTCGGTCCCTTGGCAACC 369
Db      213 TCAGGCACAGACGGCGGCTCGCTACGGTATCAACAGCGGCTTCGGACTCTTCCACAG 272
Qy      370 GCCTGATCTCAGGTGAGAATGTCCGAACGCTGCAGGCCAAATCTTGTTCATCATCTGGCCA 429
```

```

Db      273 TCATCATCGCCCGGAGAGCTCACGGAGCTGAGGAGAACCTCATCCGATCGCACTCGT 332
Qy      430 GCGCGCTGGGACCGGTGCTTGAATGCTGACGACGACGGCGCGCCATGTTCTTGGCGGCTGG 489
Db      333 CGGGACGGCGGAACCGCTGTCCCGCAGCAGACCCCGCATGTGCTCGCACTCGGCATCA 392
Qy      490 TGTGATCGCTCAGGAGCGCTCCGGTGCACAGCGAGGGGACCATCGCTCGCCTGATCGACC 549
Db      393 ACGTGTGGCCAGAGGCCACTCAGGCATCGTGTGCATACGCTGGAGCAGCTCGTGACG 452
Qy      550 TGTCTAATTCGAGCTCGCTCCGGCGCTTCCAGCGCGGACCGTGGGCGCGTTCGGGTG 609
Db      453 CTTTCAACCGCGACTGCTCTGCTGGTGTGCCAGCAAGGGCACGTTCGGCGCTCGGGAG 512
Qy      610 ACCTGACACCGCTTGGCATATGTTGCTCTGCCCTCCAGGGCCGGGGAGACTTCTCTGACC 669
Db      513 ATCTGGCCCCGCTCGCTCACTCGCGCTTGGGATGATGGGCGAAGGCCCATGTGGGACA 572
Qy      670 GGGACGGGACGCGGCTTGACGGCGCAGAGGGGCTCCGGCGCGACGCGCTGCAACCGCTCG 729
Db      573 AGTTGGCGAGGAGTTCGTTCATCAGCG-AGGCGTCCAGGTGCTGGCCAGACAGCCTA 631
Qy      730 ATCTCTCCATCCGATGCACTGGCGCTGTCAACGGGACCTCCGCCATGACCGGGATCG 789
Db      632 AGCCCGTGCAGCTCGGTGCCAAGGAAGTCTGGCCATGATCAACGGCACGCGCAGCTCATCA 691
Qy      790 CGCTGGTGAATGCTACGCGCTCGCGCATCTCGGCACTCGGGCG 833
Db      692 CGTCTGCTGGTGTGAGGCGGTCTCCGCGCTGAGAACTGTCGCG 735

RESULT 11
DN640249/c
LOCUS   DN640249                788 bp      mRNA      linear      EST 28-MAR-2005
DEFINITION UMC-bend 0A01-030-f10 Uterus (endometrium) bend Bos taurus cDNA 3',
           mRNA sequence.
ACCESSION DN640249
VERSION   DN640249.1 GI:61945444
KEYWORDS  EST.
SOURCE    Bos taurus (cow)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
           Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 788)
AUTHORS   Prather,R.S., Antoniou,E., Garverick,H.A., Green,J.A., Lucy,M.C.,
           Roberts,R.M., Smith,M.F. and Youngquist,R.S.
TITLE     USDA Grant NRI-2002-03476: Bovine ESTs: Focus on Female
           Reproduction
JOURNAL   Unpublished (2002)
COMMENT   Contact: DNA Core Facility (Bovine Project)
           Animal Science - RS Prather
           University of Missouri-Columbia
           M616 Medical Sciences Bldg., Columbia, MO 65212, USA
           Tel: (573)882-0428
           Fax: (573)884-5552
           Email: bovine@net.missouri.edu
           POLYA=No.
Location/Qualifiers
FEATURES
source
1..788
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone_lib="bend"
```

/notes=Funding: The production of ESTs submitted in this project was funded by USDA Grant MRI-2002-03476 entitled 'Bovine ESTs: Focus on Female Reproduction' to RS Prather, E Antoniou, HA Garverick, JA Green, MC Lucy, RM Roberts, MF Smith and RS Youngquist. Genetic Source: Heifers for the project were purchased from Circle A Ranch, Iberia, MO (<http://www.circlea ranch.com/home.html>). these heifers, while not registered have known Angus pedigrees going back

at least 4 generations. Samples collected: The samples consisted of the following: Germinal vesicle-stage oocytes; in vitro derived embryos (2-cell, morula, blastocyst and nuclear transfer blastocyst); in vivo blastocysts and conceptuses (days 8, 14, 16 and 18); corpora lutea (days 3, 5, 8, 14, 16, 18 and 35); ovarian follicles (days 0, non-recruited, recruited, early selected and preovulatory); oviduct (days 0, 3 and 5); endometrium (days 5, 8, 14, 16, 18 and 35); and placenta/embryo from day 35 conceptuses. Expanded descriptions of how the tissues were collected can be found at the following URL:
<http://genome.rnet.missouri.edu/Bovine/Methods.html>
 Library construction (Standard Protocol): All procedures have been described in detail elsewhere (Soares et al., 1994; Bonaldo et al., 1996; Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and the poly(A)+ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer's instructions. The libraries were constructed by E. Ferguson and R. Woods essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mcg of poly(A)+ RNA was annealed at 37 degrees with 10mcg of NotI-tag-dT18 oligonucleotide (GCGCTCGCGCGCGC-tag-T18) and reverse transcribed at c37 degrees with Superscript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (<http://genome.uiowa.edu/pubsoft/software.html>) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs were ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs were size selected by passage through cDNA size fractionation columns (Invitrogen-Life Technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pCMV-SPORT6 vector (Invitrogen). After ligation of the inserts, the plasmids were electroporated into DH10B bacteria. Library Construction (PCR Protocol): The amount of mRNA that was recovered from oocytes and embryos was quite limiting and was not sufficient for library production with the standard protocol. Therefore, PCR-based protocol was utilized for producing libraries from sources in which the amount of extracted mRNA was small (oocytes and embryos). Poly-A RNA was isolated by using the MicroPoly(A) Pure kit from Ambion (cat. # 1918). The mRNA was reverse transcribed with a NotI-tag-dT18 oligonucleotide and a SMART oligonucleotide (Clontech) modified to contain a SalI site to generate full-length cDNA with a sequence complementary to the SMART oligonucleotide. Sequences within the SMART and dT oligonucleotides were used as primers to amplify the cDNAs by PCR with pfu turbo polymerase (Stratagene). The resulting PCR products were purified, digested with NotI and SalI and size fractionated by using Chroma Spin-1000 columns (Clontech). Purified cDNA from each PCR reaction was quantitated and mixed on an equimolar basis for ligation into the pCMV-SPORT6 vector. Preliminary Library Characterization: Randomly chosen clones from each library were analyzed by restriction digestion to determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality (e.g. the presence of short polyA+ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.) and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. After production of

the libraries, equal numbers of recombinants from each library were pooled to produce a single mixed library (mega-library) for more extensive sequencing. Bioinformatics work was performed by GK Springer's Bioinformatics group (Spollen WG, Topinka CM, Khambati AA) in Computer Science at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: bovine@net.missouri.edu. Bonaldo MF, Lennon G, Soares MB. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res. 1996; 6:791-806. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Machialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(da) tails. Biotechniques 31:38-42. Soares MB, MF Bonaldo, P Jelene, L Su, L Lawton, A Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232. TAG_TISSUE=Uterus (endometrium)
 TAG_SEQ=Not found"

ORIGIN

	Query Match	5.0%	Score 92;	DB 8;	Length 788;
	Best Local Similarity	47.8%	Pred. No. 3.1e-09;	Mismatches 325;	Indels 3;
	Matches 300;	Conservative 0;			Gaps 1;
Qy	313	TCCCGAGGCGCGCATGCTTACGGAGCTGACAAACGGCTTCGGTCCCTTGGGAACCGCC	372		
Db	628	TGCGGAGAACCGCACCGCTACGGCATCAACCGGTTTCGGGCTGTGGGCTCGACTC	569		
Qy	373	TGATCTCAGGTGAGAAATGTCGGAACGGTCGAGGCGCAATCTTTGTCCATCATCTGGCCACGG	432		
Db	568	GCATCGCGCAGGAAGACCTGGAAAACCTTCAGCGCTTCCCTGGTGTGTCCACGCGCGTG	509		
Qy	433	GCGTGGGACCGGTCTTGACTGGACGACGGCGCGCCATGTTCTTGGCGGTCTGGGT	492		
Db	508	GCGTGGGTGAGCCCATCAGCATCGCTGTGTGGCGGTGTTGTTGTTCAAGGTCAACA	449		
Qy	493	CGATCGCTCAGGGAGCCCTCCGGTCCAGCAGGGGAGCCATCGCTCGCCTGATCGACCTGC	552		
Db	448	GCCTGAGCGGTGTTCTCCGGGATTCGCGGCAAGTTCATCGACGGCTGATCGCCCTGA	389		
Qy	553	TCAATTCGAGCTCGCTCGGCGCTTCCAGCGCGCGGACCGTGGGCGCGTGGGTGACC	612		
Db	388	TCAACGCGGAGGTGTACCCGCACATTCGTTGAAAGGTTTCGGTGGTGGTCCCTCCGCTGACT	329		
Qy	613	TGACACCGCTTGCGCATATGTTGCTCTCCCTCCAGGGCGGGGAGACTTCTCTGGACCGGG	672		
Db	328	TGGCGCCATTGGCCCAATGTCGCTGTGTGTGTGGGCGAAGGCAAGGCCCGTTAC---	272		
Qy	673	ACGGGACCGCGCTTGACGGGCGAAGGGCTCCGGCGCGGACGCTGCAACCGCTCGATC	732		
Db	271	AAGGCGAATGGCTGGGAGCACACGAGGGCTGAAAGTTGCGGCTGACGGCGCTGACCC	212		
Qy	733	TCTCCCATCGGATGCACTGGCGCTGTGTCAACGGGACCTCGCCATGACCGGGATCGCGC	792		
Db	211	TGGCGCCAAAGAAGGCTTGGCCCTGCTCAACGGCACTCAGGTGTCCACCGCTTATGCC	152		
Qy	793	TGGTGAATGTCAGCCCTGCGGCCATCTCGGCAACTGGGCGGTGGCGTTGACGGCCCTGC	852		
Db	151	TGCGCGCCCTGTTTCCGAAGGGCGAAGACCTGTTGCGCGGTGCAATTTGGCTGTGGCGCCCTGA	92		
Qy	853	TTGCGGAATGTTCTGAGAGGCGGACCGAGGCAATGGGCGCGGCACTGTCCGACCTTCCGGC	912		
Db	91	CGGTGGGAAGCGGTACTGGTTTGGCGCTCGCGCTTCGACGCGACGCAATTCACGCTGCCCGT	32		
Qy	913	CGCATCCCGACAGAGAGCGCCGACG	940		
Db	31	GCCAGCGTGGCCAGATCGACTCTGCGGC	4		

RESULT 12

CX949440/c
 LOCUS UMC-bcl.0A02-030-b05 Corpus Luteum (CL) bcl Bos taurus cDNA 3',
 DEFINITION mRNA sequence.
 ACCESSION CX949440
 VERSION CX949440.1 GI:58762967
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 930)
 AUTHORS Prather,R.S., Antoniou,E., Garverick,H.A., Green,J.A., Lucy,M.C.,
 Roberts,R.M., Smith,M.F. and Youngquist,R.S.
 TITLE USDA Grant NRI-2002-03476: Bovine ESTs: Focus on Female
 Reproduction
 JOURNAL Unpublished (2002)
 COMMENT Contact: DNA Core Facility (Bovine Project)
 Animal Science - RS Prather
 University of Missouri-Columbia
 M616 Medical Sciences Bldg., Columbia, MO 65212, USA
 Tel: (573)882-0428
 Fax: (573)884-3552
 Email: bovine@net.missouri.edu
 POLYA=No.

FEATURES
 source
 1..930
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /clone_lib="bcl"
 /notes="Funding: The production of ESTs submitted in this
 project was funded by USDA Grant NRI-2002-03476 entitled
 'Bovine ESTs: Focus on Female Reproduction' to RS
 Prather (Primary Investigator), E Antoniou, HA Garverick,
 JA Green, MC Lucy, RM Roberts, MF Smith and RS Youngquist.
 Genetic Source: Heifers for the project were purchased
 from Circle A Ranch, Iberia, MO
 (http://www.circlearanch.com/home.html). These heifers,
 while not registered have known Angus pedigrees going back
 at least 4 generations. Samples collected: The samples
 consisted of the following: germinal vesicle-stage
 oocytes; in vitro derived embryos (2-cell, morula,
 blastocyst and nuclear transfer blastocyst); in vivo
 blastocysts and conceptuses (days 8, 14, 16 and 18);
 corpora lutea (days 3, 5, 8, 14, 16, 18 and 35); ovarian
 follicles (days 0, non-recruited, recruited, early
 selected and preovulatory); oviduct (days 0, 3 and 5);
 endometrium (days 5, 8, 14, 16, 18 and 35); and
 placenta/embryo from day 35 conceptuses. Expanded
 descriptions of how the tissues were collected can be
 found at the following URL:
 http://genome.rnet.missouri.edu/Bovine/Methods.html.
 Library construction (Standard Protocol): All procedures
 have been described in detail elsewhere (Soares et al.,
 1994; Bonaldo et al., 1996; Jiang et al., 2001). Total
 cellular RNA from each sample was isolated by using
 STAT-60 reagent (Tel-Test, Friendswood, TX) and the
 poly(A)+ RNA was obtained by two rounds of purification
 with the Oligotex mRNA isolation kit (Qiagen) according to
 the manufacturer's instructions. The libraries were
 constructed essentially as described by the manufacturer's
 instructions provided with the SuperScript plasmid System
 (Invitrogen, cat. no. 18248-013). Briefly, 1mcg of
 poly(A)+ RNA was annealed at c37 degrees with 10mcg of
 NotI-tag-dT18 oligonucleotide (GCTGTCGGCGCG-tag-T18)
 and reverse transcribed at c37 degrees with SuperScript II
 (Invitrogen) reverse transcriptase (Jiang et al., 2001).
 The 'tag' represents a tissue/stage-specific ten-base
 sequence identifier
 (http://genome.uiowa.edu/pubsoft/software.html) present in
 the oligonucleotide used to prime first-strand synthesis.

Second strand synthesis was performed with T4 DNA
 polymerase in the presence of DNA ligase and RNase H.
 After second strand synthesis, the double-stranded cDNAs
 were ligated to SalI adapters (Invitrogen-Life
 Technologies) and digested with NotI. The cDNAs were size
 selected by passage through cDNA size fractionation
 columns (Invitrogen-Life Technologies). The cDNAs derived
 from each developmental stage of a particular tissue were
 mixed on an equimolar basis and ligated directionally into
 the NotI and SalI sites of the pCMV-SPORT6 vector
 (Invitrogen). After ligation of the inserts, the plasmids
 were electroporated into DH10B bacteria. Library
 construction (PCR Protocol): The amount of mRNA that was
 recovered from oocytes and embryos was quite limiting and
 was not sufficient for library production with the
 standard protocol. Therefore, PCR-based protocol was
 utilized for producing libraries from sources in which the
 amount of extracted mRNA was small (oocytes and embryos).
 Poly-A RNA was isolated by using the MicroPoly(A) Pure kit
 from Ambion (cat. # 1918). The mRNA was reverse
 transcribed with a NotI-tag-dT18 oligonucleotide and a
 SMART oligonucleotide (Clontech) modified to contain a
 SalI site to generate full-length cDNA with a sequence
 complementary to the SMART oligonucleotide. Sequences
 within the SMART and dT oligonucleotides were used as
 primers to amplify the cDNAs by PCR with pfu turbo
 polymerase (Stratagene). The resulting PCR products were
 purified, digested with NotI and SalI and size
 fractionated by using Chroma Spin-1000 columns (Clontech).
 Purified cDNA from each PCR reaction was quantitated and
 mixed on an equimolar basis for ligation into the pCMV-
 SPORT6 vector. Preliminary Library Characterization:
 Randomly chosen clones from each library were analyzed by
 restriction digestion to determine average insert size (96
 clones) and by sequencing (~4 96-well plates) to confirm
 library quality [e.g. the presence of short polyA+ tails,
 genomic DNA contamination (must be <1%), ribosomal RNA
 clones (must be <1%), etc.] and to provide a sequence
 database representing the predominant clones in each
 library. The clones were sequenced at the University of
 Missouri-Columbia DNA Core Facility. After production of
 the libraries, equal numbers of recombinants from each
 library were pooled to produce a single mixed library
 (mega-library) for more extensive sequencing.
 Bioinformatics work was performed by GK Springer's
 Bioinformatics group in Computer Science at the University
 of Missouri-Columbia. Clone Requests: Requests for clones
 should be made to the Director of the University of
 Missouri DNA Core facility at: bovine@net.missouri.edu.
 Bonaldo MF, Lennon G, Soares MB. Normalization and
 Subtraction: Two approaches to facilitate gene discovery.
 Genome Res, 1996; 6:791-806. Jiang H, Bivens NJ, Ries JE,
 Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion
 BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing
 cDNA libraries with fewer clones that contain long
 poly(dA) tails. Biotechniques 31:38-42. Soares MB, MF
 Bonaldo, P Jelene, L Su, L Lawton, A Efstratiadis. 1994.
 Construction and characterization of a normalized cDNA
 library. Proc Natl Acad Sci, 91:9228-9232.
 TAG_TISSUE=Corpus Luteum (CL)
 TAG_SEQ=Not found"

ORIGIN

Query Match 5.0%; Score 92; DB 8; Length 930;
 Best Local Similarity 47.8%; Pred.No. 3.le-09;
 Matches 300; Conservative 0; Mismatches 325; Indels 3; Gaps 1;
 QY 313 TC GCGAGGCGCGCCATGTCTACGGACTGACACACCGGCTTCGGTCCCTTCGGAACGCC 372
 |||||
 Db 628 TC GCGAGAACCGACCGCTACGGCATCAACACCGGCTTCGGCCTTCGCACTC 569
 |||||
 QY 373 TGATCTCAGTGAGATGTCCGACCGTCGAGCCCAATCTTGTCCATCATCTTGGCAGCG 432
 |||||

```
Db GCATCGCCAGCGAGACCTTGAACACCTTACGGTTCCCTGGTGTGTGTCCTCCACGGCGGTG 509
Qy GGTGGGACCGGTGTTGATCGAGACGCGCGCGCCATGTTCTGGCGGTCTGGTGT 492
Db GGTGGGTGAGCCCATCAGCGATGCGGTGTGTGTCATGTTGTCTCAAGGTCAACA 449
Qy CGATCGCTCAGGAGCTTCCGGTGCAGCGAGGGGACCATGCTCGCTGTATCGACCTGC 552
Db GCCTGAGCGGTGTTCTCCGGGATTCGCGGACAGTCTATCAGCGGCTGATCGCCCTGA 389
Qy TCAATTCGAGCTCGCTCCGGCGTTTCCAGCGCGGCGACGCTGGCGCGGTGGGTGACC 612
Db TCAACGCGGAGGTACCGCACATTCGCTTGAAGGTTTCGGTGGGTGCTCCGCTGACT 329
Qy TGACACCGCTTGGCATATGCTGCTGCTCCAGGCGGGGAGACTTCTTGGAGCGGG 672
Db TGGCGCCATTGGCCACATGTCGCTGGTGTCTGCTGGGGAAGGCAAGGCCGCTTAC--A 272
Qy ACGGACCGCGCTTGACGCGCGAGAGGCTCCGGCGCGAGCGCTGCAACCGCTCGATC 732
Db AAGCGAATGCTGGAGCCACCGAGGCGCTGAAAGTTGCGGCTGAGCGCGCTGACCC 212
Qy TCTCCATCGGATGCACTGGCGCTGGTCAACGGGACCTCGGCCATGACCGGGATCGGCG 792
Db TGGCGCCAAAGAGGCTTGGCCCTGCTCAACGGGCACTCAGGTGTCACCGCTTATGCC 152
Qy TGGTGAATGCTACCGCTGCGCCATCTCGGCAACTGGCGGTGCGGTGAGCGCCCTGC 852
Db TGGCGGCGCTTTCGAAGCGAAGACCTGTTGCGCGGTGTCATTTGGCTGTGGCGCCCTGA 92
Qy TTGGGGAATGCTGAGAGCGGAGGAGGATGGCGCGGCGGCACTGTCGACCTGCGGC 912
Db CGGTGAAGCGGATCTGGGTTCGCGCTCGCGCTTCGACGCGACGATTCACGCTGCCCGTG 32
Qy CGCATCCCGGACAGAAAGCGCGCGAGC 940
Db GCCAGCGTGCCAGATGCACTTGGCG 4
```

```
RESULT 13
DR785713
LOCUS DR785713 847 bp mRNA linear EST 27-JUL-2005
DEFINITION ZM_BFBF Zea mays cDNA 5', mRNA sequence.
ACCESSION DR785713
VERSION DR785713.1 GI:71299412
KEYWORDS EST.
```

```
SOURCE
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 847)
Kim, H., Collura, K., Wisotski, M., Smart, D., Kudrna, D., Muller, C.,
Rao, K., Haller, K., Wing, R., Soderlund, C., Walbot, V. and Yu, Y.
Maize Full-length cDNA Project
Unpublished (2005)
Contact: Yeisoo Yu
Arizona Genomics Institute
The University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9585
Fax: 520 621 1259
Email: yeisoo@genome.arizona.edu
Plate: 0002 row: F column: 14.
```

```
FEATURES
source
1. 847
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/tissue_type="mixed (silks, husks, ears, pollen, shoot
tips, leaf, root tips, whole seed, embryo)"
/dev_stages="varies by tissue"
```

```
/lab_host="DH10B T1 phage resistant"
/clone_lib="ZM_BFB"
/notes="Vector: pCMV-SPORT 6.1; Site 1: EcoRV; Site 2:
NotI; Maize Full length cDNA library (3530 library)
created by Invitrogen from multiple tissues; Organ: silks,
husks, ears, pollen, shoot tips, leaf, root tips, whole
seed, embryo. This is a Gateway compatible vector, for
permitting clone movement to new vector backbones, for
expression in diverse host cells using recombination
rather than restriction enzymes. poly(A)+ mRNA was
prepared by Invitrogen, and equimolar amounts of RNA from
each of the 12 tissue samples were mixed together for
a normalization step was conducted against the mixture of
RNA sources. tissues prepared: 1. just emerging silks; 2.
immature husks from ears of sample #1; 3. 20 day aleurone; 4.
immature tassels, stages from 1-2 mm to 1-2 cm; 5. 2 mm to
2 cm ears; 6. pollen; 7. 1 cm vegetative shoot tips from
15 day old seedlings; all leaves with an expanded or
partially expanded sheath were removed; 8. mature leaf
tissue; 9. 0.5 cm long root tips from 15 day old
seedlings; 10. 10 day whole seed; 11. 12 day endosperm and
embryo; 12. 17 day endosperm and embryo. All of the
sequenced clones in Maize Full-length cDNA Project will be
archived at the University of Arizona. Clones, high
density filters and amplified library can be ordered from
the University of Arizona
(http://www.genome.arizona.edu/orders/)."
```

ORIGIN

```
Query Match 4.9%; Score 90.6; DB 8; Length 847;
Best Local Similarity 49.5%; Pred. No. 6.2e-09;
Matches 361; Conservative 0; Mismatches 344; Indels 24; Gaps 4;
Qy 266 GGCGCGGACCGGTGCGTCCGAGCGCGGTCCGCTGTCATCCGAGCGCGG 325
Db |||||
Qy 25 GGCGCGCGCGGTCAAGGCGAGGAGTGTCTGACTCATGCGCCACGCGG 84
Db |||||
Qy 326 CCATGTCTACGAGTGAACAACCGGTTCGGTCCCTTCGGAACCGCTGATC----- 377
Db |||||
Qy 85 CGACATCTACGGCTCAACACCGGCTTCGGCGGCACTCCACCGCGCACCAAGGACGG 144
Db |||||
Qy 378 TCAGTGAATGTCGAGACGCTGCGGCGCAATCTTGTCCATCATCT-----GGCCAGCGG 433
Db |||||
Qy 145 GCCCGCGCTCCAGTTCGAGTGTCTCAGGCATCTCAACCGCGGAATCTTCGCGACCGGCGAG 204
Db |||||
Qy 434 CGTGGACCGGTGCTTGACTGGACGACGCGCGGCGCATGGTCTGGCGCGTCTGGTGTGTC 493
Db |||||
Qy 205 CGACGGGACACGCTGCGGTGCGAGGTACCGCGCGCGGATGCTGGTGGCATCAACAC 264
Qy 494 GATCGTCAAGGAGCGCTCCGGTGCAGGAGGGACCATCGCTCGCTGTGATCGACTGCT 553
Db |||||
Qy 265 CCTCTCCAGGGTACTCCGGCATCCGCTTCAGATCTCTCGAGGCGCATCAGAAAGTGTCT 324
Qy 554 CAATTCCGAGTCTCGCTCCGGCGGTTCCAGCGCGGCGGACGCTGGCGGCTGGGTGACTCT 613
Db |||||
Qy 325 CAACACCGGTGTAGCCCTGCTGCGGCGCACCATCACCGGCTCGGGCGACTCT 384
Db |||||
Qy 614 GACACCGCTTGGCATATGGTGTCTGTCTCCAGCGCGGCGGAGACTTCTCTGG-----A 667
Db |||||
Qy 385 GGTCCCGCTCTCTACATCGCGCGCTCATCAGCGGCGCGCCCAACGCGGAGCGCGCTAC 444
Qy 668 CCGGGAAGGAGCGGCTTGAAGCGGAGAGGCTCCGGCGCGGAGCGGCTGCAAC--- 724
Db |||||
Qy 445 CGTCGAGGAGGAGGTGGAGCGCGCGGCGGTTCAGATCGCGGCGCATCGAGGGCGG 504
Qy 725 ---GCTGATCTCTCCATCGGATGCACTGCGCGCTGTTCAACGGGAGCTCGGCATGAC 781
Db |||||
Qy 505 CTTCTTCAAGCTCAACCCCAAGGAGGCGCTCGCATGCTCAACGGGACGCTCGTGGGCTC 564
Qy 782 CGGGATCGCGTGTGATGCTCAGCGCTGCGGCGCATCTCGGCAACTCGGCGGTGGCGCTT 841
Db |||||
Qy 565 CGCGCTCGCGGCCACCGGTGATGTACGAGCCCAACGTCCTCGCGCGTCTGTGTGAGGAGTCT 624
```


Tel: 520 626 3967
Fax: 520 621 9288
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 03 row: A column: 03
Seq primer: gta aaa cga cgg cca gtg.
Location/Qualifiers
1..866
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB03A03"
/issue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

ORIGIN

Query Match
Best Local Similarity 4.9%; Score 90.4; DB 6; Length 866;
Matches 389; Conservative 0; Mismatches 376; Indels 33; Gaps 4;

Qy 430 GCGCGTGGGACCGGTGTTGACGACGCGCGCGCCATGTTCTGGCGGTCTTGG 489
Db 54 GCTCGATGGCCACACGCTGCGGTCGGAGACGGTGGGGGGCCATGCTCGTGGCATCA 113
Qy 490 TGTGATCGCTCAGGAGCCTCCGGTGCACGAGGGGACCATCGCTCGCCTGATCGACC 549
Db 114 ACACCCCTCTCCAGGGCTACTCCGGCATCCGGTTCGAGATCTCGAGGCCATCACCAAGC 173
Qy 550 TGCTCAATTCGGAGCTGCTCCGCGGTTCCAGCGCGGACCGTGGCGCGCTCGGTG 609
Db 174 TGCTCAACACCGGGCTACGCGGCTGCTCGCGCTCGGTGGGACCATCACCGCGTCCGGTG 233
Qy 610 ACTTGACACCGCTTGGCATATGGTGTCTGCTCTCAGGGCGGGGAGACTTCTTGGAC- 668
Db 234 ACCTGGTTCCCTGTCTTACATTGCGGCTCTCATCACCGCGCGCCCAACCGCAGGCCA 293
Qy 669 -----CGGAGCGGACGGGCTTGAACGCGCAGAAAGGCTCCGGCGGAGACGGCTGCA-- 721
Db 294 TCTCGCCGACGCGGAGGAGTGGACGCGCGAGGGGTTCAAGCTGCGCGCATCGAGG 353
Qy 722 ----ACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGCTCAACGGGACCTCCGCCA 777
Db 354 GTGGCTTCTTCAACGCTGAACCCCAAGGAAGGTCTCGCCATCGTCAATGGCACGTCCTGG 413
Qy 778 TGACCGGATCGCGCTGGTGAATGCTCAACGCTGCGGCATCTCGGCAACTGGGCGGTGG 837
Db 414 GGTGGCGCTCGCGGCCACCGGTGATGTTGACGCGCAACATCTCGCGTCTGTCCGAGG 473
Qy 838 GGTGACGGCCCTGCTTGGCAATGTCTGAGAGCCGGAACGAGGCATGGGCCCGGCGAC 897
Db 474 TGCTCTCGGCGGTGTTGCGAGGTGATGAACGCGAAGCGGAGTACACCGACCACTGA 533
Qy 898 TGTCCGACCTGGCGCGCATCCCGGACAGAGAGCGCGGACGAGGCTGGCGCCCGCG 957
Db 534 CCCAAGAGCTGAAGCACACCCCTGGGTGATCGAGGGCGCGCCATCATGGAGCACATCC 593
Qy 958 TGACGCGCAGCGCGGCTGGTGGTGGACGCTATTGCGAGCGGAGGCTCGACCGCGCGG 1017
Db 594 TCGCCGGGAGTCTGTTCAAGAGCCACGCCAAGA-----AGGTGAACG 635
Qy 1018 ATATCGGGAACGAGCGGAGCGGGGAGGATGCTACAGCTGCTGCTGCGCTCCGAGG 1077
Db 636 AGATGGACCCGCTGCTGAAGCCGAAGCAGGACAGGTACGCGCTCCGACAGCTCGCGCAGT 695
Qy 1078 TTCTCGGGCGGGCTTGCACACGCTCGCATGGCATGCCGGTGTGACGATCGAGCTGA 1137

Db 696 GGCTCGGCCCGCAGATCGAGGTCAATCCGCGCGCGCCACCAAGTCCATCGAGCGCGAGGTCA 755
Qy 1138 ACGCGGTGACCGACAATCCGGTGTTCGCGCCGATGGCAGCGTGCCTCCCTGCAACGGGG 1197
Db 756 ACTCGGTGAACGACAACCGCGTGTGATCGACGTCCACCGCGGCAAG---GCGCTCCACGGGG 812
Qy 1198 GCAATTTTCATGGGCGAGC 1215
Db 813 GCAACTTTCAGGGGCAAC 830

Search completed: December 10, 2005, 17:16:23
Job time : 7544 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 02:01:09 ; Search time 343 Seconds
(without alignments)
9587.432 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccagcgccatccacgag.....tcagctcgatcccgagagg 1850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA.*
- 1: /cgm2_6/ptodata/1/ina/1 COMB.seq.*
 - 2: /cgm2_6/ptodata/1/ina/5 COMB.seq.*
 - 3: /cgm2_6/ptodata/1/ina/6A COMB.seq.*
 - 4: /cgm2_6/ptodata/1/ina/6B COMB.seq.*
 - 5: /cgm2_6/ptodata/1/ina/H COMB.seq.*
 - 6: /cgm2_6/ptodata/1/ina/PCITUS COMB.seq.*
 - 7: /cgm2_6/ptodata/1/ina/PP COMB.seq.*
 - 8: /cgm2_6/ptodata/1/ina/RE COMB.seq.*
 - 9: /cgm2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	215.4	11.6	1527	3	US-09-902-540-8003
2	215.4	11.6	6439	3	US-09-902-540-813
3	171.8	9.3	1548	3	US-09-902-540-813
4	171.8	9.3	2295	3	US-09-252-991A-16475
5	163.2	8.8	1527	3	US-09-252-991A-16126
6	161.8	8.7	1542	3	US-09-489-039A-6404
7	153.8	8.3	1536	3	US-09-833-745-12
8	125.4	6.8	1557	3	US-09-252-991A-16124
9	125.4	6.8	1563	3	US-09-252-991A-16478
10	110.4	6.0	2787	3	US-09-624-693A-16
11	110.4	5.9	963	3	US-09-252-991A-16368
12	106.6	5.8	2475	3	US-09-624-693A-20
13	103.2	5.6	2061	3	US-09-765-873A-31
14	103.2	5.6	2151	3	US-09-627-216A-7
15	103.2	5.6	2151	3	US-09-765-873A-7
16	101.6	5.5	2151	3	US-09-627-216A-9
17	101.6	5.5	2151	3	US-09-765-873A-9
18	101.6	5.5	2439	3	US-09-624-693A-18
19	100.6	5.4	2476	3	US-09-221-017B-876
20	88.4	4.8	798	3	US-09-252-991A-16232
21	83	4.5	2419	3	US-09-624-693A-12
22	78.4	4.2	3139	3	US-09-949-016-303
23	78.4	4.2	3786	3	US-09-919-039-351
24	76.8	4.2	3141	3	US-09-949-016-4130

25	67.6	3.7	2465	3	US-10-439-479-24	Sequence 24, Appl
26	66.2	3.6	438	3	US-09-252-991A-16256	Sequence 16256, A
27	63.8	3.4	1455	3	US-09-615-192A-245	Sequence 245, App
28	57.8	3.1	1320	3	US-09-724-797-85	Sequence 85, Appl
c 29	57.4	3.1	471	3	US-09-252-991A-16230	Sequence 16230, A
30	57.2	3.1	29272	3	US-09-902-540-1217	Sequence 1217, Ap
31	54.6	3.0	94750	3	US-09-596-002-38	Sequence 38, Appl
32	54.4	2.9	1512	3	US-09-902-540-4044	Sequence 4044, Ap
c 33	54.4	2.9	24986	3	US-09-902-540-1200	Sequence 1200, Ap
c 34	54	2.9	3354	3	US-09-902-540-9590	Sequence 9590, Ap
c 35	54	2.9	13346	3	US-09-902-540-1089	Sequence 1089, Ap
36	53.6	2.9	77536	3	US-09-410-551B-1	Sequence 1, Appli
37	53.6	2.9	77536	3	US-09-940-316B-1	Sequence 1, Appli
38	53.4	2.9	1500	3	US-09-252-991A-6254	Sequence 6254, Ap
c 39	53.4	2.9	1695	3	US-09-252-991A-6044	Sequence 6044, Ap
c 40	53	2.9	2561	3	US-09-616-289-48	Sequence 48, Appl
c 41	53	2.9	2561	3	US-09-976-740-48	Sequence 48, Appl
c 42	53	2.9	6360	3	US-10-212-962-2	Sequence 2, Appli
43	52.4	2.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
44	52.4	2.8	44377	2	US-08-804-198-1	Sequence 1, Appli
45	51.8	2.8	774	3	US-09-252-991A-15298	Sequence 15298, A

ALIGNMENTS

RESULT 1
US-09-902-540-8003
; Sequence 8003, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8003
; LENGTH: 1527
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8003

Query Match		11.6%	Score 215.4; DB 3; Length 1527;
Best Local Similarity		50.0%	Pred. No. 8e-34;
Matches 753; Conservative		0;	Mismatches 691; Indels 63; Gaps 6;
Qy	203	CATCGATCTGGACCGAGCCCATCGGTGGCGAGCGCGGGGATTGTCCTTGCCCC	262
Db	36	CCTGAAGCTCGAGGAATCTCCAGGTGGTTCGCAACGAGGCCACCGTGAGCTGCCCC	95
Qy	263	TCGCGCGCGAGCGGTGCGTCCGAGCGCGGTGCGCGCTGTCTATCCGCGAGGC	322
Db	96	CGAGCGCGCACCGCGTGGCGCTCGCGCGCTGGTGACCGCTCGCGCGGAGA	155
Qy	323	CGCGCATGTCTACGGACTGACACCGGCTTGGTCCCTTCGAGACCGCTGATCTCAGG	382
Db	156	CAGCGCGCTTCAGGCATCAACCGGCTTGGCACGTTGGCGAGTTCGCATCGACAA	215
Qy	383	TGAGATGTCGAGACGCTGCAGGCAATCTGTCCATCATCTGCGCAGCGCGGGAGCC	442
Db	216	GAGGACCTGGCGACCTCCAGGCAACCTATCTCTCCACGCGTGTGGCGTGGCAC	275
Qy	443	GGTGTCTTACTGAGCAGCGCGCGCCATGTTGTCGCGCGTGTGTTGATCGTCTCA	502
Db	276	GCCCTCCCTTCGGAAGCGGGCGCTCTGCTGCTCGCTGCAACGCTGCTCGGCA	335
Qy	503	GGGAGCCTCGGTGCCAGCGAGGGGACCATGCTCGCTGATCGACTGCTCAATTCCGA	562

Db 336 GGGCTACTCCGGCATCCGCATGGAGACGCTGGCCCTGGACATGCTGAACCGGA 395
Qy 563 GCTCGCTCCGCGCGTTCACAGCGCGGACGCTGGGGCGCTCGGGTGACCTGACACCGCT 622
Db 396 GCTCGTCCGCGTGTCCCGAGCGGGGACGCTGGGGCGCTCCGGGATCTCGCCCCGCT 455
Qy 623 TGGGCATATGCTGCTCTGCCCTCCAGGCGCGGGGAGACTTCTGGACCGGACCGGACGCG 682
Db 456 GCGGCACCTCGCGCTGCTCTTCATCGGGAAGGTGAAGCCTTCTATCAGGGCCA--GCG 512
Qy 683 GCTTGACGGCGCAGAAGGGTTCGCGCGCGGACGCGCTGCAACCGCTGATCTTCCCATCG 742
Db 513 GATGCCCGCAGCAGCGCTGGAGCGCGCGCTGCAACCGGTGTGTGGAGGCCAA 572
Qy 743 CGATGCACTGGCGCTGCTCAACGGGACCTCGGCATGACCGGATCGCGCTGGTGAATG 802
Db 573 GGAGGGGCTCGCCCTGGTGAACGGCACACAGGGCCATGTGCGCGTGGGACCCCTGTCCA 632
Qy 803 TCACGCGCTGCGGCATCTCGGCAACTGGGCGGTGGCGTTGACGCGCCTGCTTGGGAATG 862
Db 633 GCTTCGGCGGAGTCCCTGGCGGACATCGCGGAGTGGCGGGGCCATGACGCTGAGGG 692
Qy 863 TCTGAGAGCGCGACCGAGGATGGGCGCGGCACTGTTCGACCTGGCGCGCATCCCGG 922
Db 693 GCTGCTGGGAAGCCACAAGCCCTTCAATCTGAGATTACAGACGTCGCGCGCACCGGG 752
Qy 923 ACAGAGGAGCGCGCAGGAGTGGCGCGCGCTGGAGCGGACGCGCGGTGGTCCG 982
Db 753 CCAGAAGGAGCTCGCGCGCACCTGGCGGCGCATCTCTGGTGGACAGCGAGCTGGTGATC 812
Qy 983 GCACGTCATTGCGGAGCGGAGGCTCGACGCGCGCGATATCGGACGAGCGGAGCGGG 1042
Db 813 GCAGTCA-----ACTGCAGCAAGT 833
Qy 1043 GCAGGATGCTACAGCTGCGTGGCTCCGCGAGTTCTCGGGCGGGCTTCGACACGCT 1102
Db 834 GCAGGACCCCTACTCCTCGCGTGCATGCGCGAGGTGACGGCGCGCGCGGAGGCGAT 893
Qy 1103 CGCATGGCATGACGGGTGTGAGCATCGAGCTGACGCGGTGACGACCAATCCGGTGT 1162
Db 894 CGCGTTCTCCCGCGCATCTCGAGGTGGAGTCAACAGCGCAGGACCAACCGCTCGT 953
Qy 1163 TCCGCCCGATGGCAGCGGTGCGCGCTGACGCGGGCAATTTATGGGCGAGCATGTGC 1222
Db 954 ---GTTTCGGGACACGAGGCGCATGCTGTCGGGCGGCACTTCCAGGCGAGCCCATCTC 1010
Qy 1223 GCTGACGCTCGATGCTGCGCACGCGCGTCAACGTTCTGGCGGGCTTTCGGAGGCGCA 1282
Db 1011 CCTGGCCATGGAAGTGTGGCGATGGCGCTGACGCAACTGTCTCATCAGCGAGCGCG 1070
Qy 1283 GATTGCACTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGG 1342
Db 1071 CGTGGAGCAGCTCGTGAACCGCTGCTGCTCAACCTGCGCGGT-----TCTCGCGAA 1124
Qy 1343 CCCCGCGGGTGAATTCGCGGTTTCAGGCGCACAGGTGACGCGGACGCGCTCTGGC 1402
Db 1125 GAATCTCGGGTTGAATCTCCGGCTTATGATCGGCGAGGTGAACAGCGCGCGCTGGTGG 1184
Qy 1403 CGAGATGCGAGCCACGCGGA---CCTGCTCGATTCATTCGATCTCAACGAAACCGCCAA 1459
Db 1185 CGAGTCCCGGTGCTGAGCCACCCCGCTCGTGGATTGATTCGTCATCCCGGCGCG 1244
Qy 1460 TCAGGATGTTGTTCTCGTTGGGACCATCGCGCGCGCTTTCGCGGAGAGATCGACCG 1519
Db 1245 AGAGGACCAAGTGTCCATGGGCAAGACGCGCGCGCTCAAGGGCGGTGAGTTCAGCACTT 1304
Qy 1520 TTGGGCGGAGATCTTTCGAGTCTCTGCTGCTGCTTTCGACAAGCTGCGGAGTGGCGTG 1579
Db 1305 CGCCCGTTGCTCGGATTGAATCTTGTGGCGGCGGAGGCTTGAATTCGCGCT 1364
Qy 1580 CGGACGCGGCTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGGTGAGGCCCTTGGCGGA 1639

Db 1365 GCGCTGAAGCCCGGCAAGGGCGCCCTCGCGCGTACGAGCTGG-----TGGCTC 1415
Qy 1640 GCAGTTCCCGCGCTTGAAGACGACCGGCCCTTGGACAGAAATTCGCGCTTGTCTAC 1699
Db 1416 GAAGTTCGCCCATCGACAGGACCGGAGCTGACCCGGAACATCGAGCGGTAGCCA 1475
Qy 1700 GCACCTC 1706
Db 1476 GCTCGTC 1482

RESULT 2
US-09-902-540-813
; Sequence 813, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 813
; LENGTH: 6439
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-813

Query Match 11.6%; Score 215.4; DB 3; Length 6439;
Best Local Similarity 50.0%; Pred. No. 9.2e-34;
Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;

Qy 203 CATCGATCTGGACACGCGCCCATGCGCTGGCGAGCGCGCGCGGATTTCTCTTGCCCC 262
Db 1355 CCTGAAGCTCGAGGAATCTCCAGGTGGCTCGCAAGAGGCCACCGTGAGCTGCGCC 1414
Qy 263 TCGCGCGCGGACCGGTGCGTGCAGAGCGCGGCTGCGGCTGTCATCCGCGAGGC 322
Db 1415 CGACGCGCGCACCGCGTGGCGCTTGGCGCGCTTGGTGAACCGGTGCGCGCGGAGA 1474
Qy 323 GCGCATGTCTAGGACTGACACCGGCTTGGTCCCTTGGCAACCGCTGATCTCAGG 382
Db 1475 CAGCCCGCTTACGCGATCAACACCGGCTTGGCAAGTTGGCGGAGTCCGCATCGACAA 1534
Qy 383 TGAGAAATGTCGAAACGCTGACGCGCAATCTTGTCCATCATCTGGCCAGCGCGCTGGGACC 442
Db 1535 GAAGGACCTGGCGGACCTCCAGCGCAACCTCATCTCTCCACGCGTGTGGCGTGGCAC 1594
Qy 443 GTGTCTTGAATGGAACGCGCGCGCCATGTTCTGGGCGGCTGTGGTGTGATTCGTCA 502
Db 1595 GCGCCCTCCCTTCCGGAAGCGCGGCGCTCTGCTGCTCGCTGCAACGCTGCTGCGCAA 1654
Qy 503 GGGAGCTCGGTGCGAGCGGAGGACCATCGCTCCGCTGATCGACCTGCTCAATTCCGA 562
Db 1655 GGGTACTCGGCATCCGATGAGAGCGTGGCGCTTGGCGCTGGAATGTAACCGGGA 1714
Qy 563 GCTCGCTCGCGCGTTCGACGCGCGGCGAGTGGGCGGTGGGTGACCTGACACCGCT 622
Db 1715 GCTGTGCGCGTGTCCCGAGCGGGGACGCTGGGCGGCTCGGGGATCTCGCCCGCT 1774
Qy 623 TCGCATATGTTGCTGTGCTCCAGGGCGGGGAGACTTCTTGGACCGGGAACGGAGCGG 682
Db 1775 GCGCACCTTGGCGTCTTCTTATCGGCGAAGGTGAAGCTTCTATCAGGGCCA--GCG 1831
Qy 683 GCTTGAACGCGCAGAGGCTCCGCGCGGAGCGGCTGCAACCGCTGATCTCTCCCATCG 742
Db 1832 GATGCCCGGAGACGAGCGCTGGAGCGCGCGCGCTTGCACACCGGTGTGTGGAGGCCAA 1891

Qy	743	CGATGCACTGGCGCTGTGTCAAACGGGACCTCCGCCATGACCGGGATCGCGTGTGTGAATGC	802
Db	1892	GGAGGGCTTCGCCCTGTGTGAACGGGCACACAGGCCATGTGCGCGTGGGACACCTGTCTCCA	1951
Qy	803	TCACGCTTGCCGCCATCTCGGCAACTGGGCGGTGGCGTTGACGGCCCTGCTTGGCGGAATG	862
Db	1952	GCTTCGCGCGGAGTCCCTTGGCGGACATCGCCGACGTGCGGCGCCATGACGCTTGAGGG	2011
Qy	863	TCTGAGAGCCCGGACCGAGGATGGGCGCGGGCACTGTCCGACTTGCGGCGGCATCCCGG	922
Db	2012	GCTGCTGGGAAGCCACAAGCCCTTCATTCTCTGAGATTACAGACTCCGCGGCACCCCGGG	2071
Qy	923	ACAGAAGAGCCCGCACCGAGGCTGCGGCCCGCGTGGACGGCAGCGCGGGGTGTCCG	982
Db	2072	CCAGAAAGACGTTCGCGCGCACCTTGCGGCGCATCTTGTTGACAGCAGCTTGTGSGATC	2131
Qy	983	GCACGTCAATTGCCGAGCGGAGGCTCGACGCCCGCGCATATCGGACGAGCCGAGGCGGG	1042
Db	2132	GCACGTCA-----ACTGCAGCAGGT	2152
Qy	1043	GCAGGATGCCCTPACAGCTTGCCTCGCTCCGACAGTTCTCGGGCGGGCTTCGACACGCT	1102
Db	2153	GCAGGACCCCTACTCTCCCTGCGTGCATGCGCAGGTGCACGGCGCGCGCGCAGGGCAT	2212
Qy	1103	CGCATGCAATGACCCGGTGTCTGACGATCGAGCTGAACCGGTGACGACACATCCGGTGT	1162
Db	2213	CGCGTTCTCCCGCGCATCTCTGGAGGTGGAGGTCAACAGCGCGACGACAAACCCGCTCGT	2272
Qy	1163	TCCGCCGATGGCAGCGTGC CGCCCTCTGCACGGGGCAATTTCAATGGGCCAGCATGTGGC	1222
Db	2273	--GTTTCGCGGACACGGAGCGCATCGTGTGCGGCGGCAACTTCCACGGCCAGCCCATCTC	2329
Qy	1223	GCTGACGTCCGATGCGTTCGGCACGGCGGTCAACGTTCTGCGGGGCTTGGGAGCGCCA	1282
Db	2330	CCTGGCCATGGACGTGTGGCGATGGCGCTGACGCAACTGTGTCTTCATCAGCGAGCGCGC	2389
Qy	1283	GATTGCACGTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCTTCCTCCACCGGGG	1342
Db	2390	CGTGGAGAGCTCGTGAACCCGTCGCTGTCCAACTGCGCGGT-----TCTTGGCGAA	2443
Qy	1343	CCCGCGCGGTGTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGACCGCGTCTGTGGC	1402
Db	2444	GAACCTCGGGTTGAACCTCCGCTTCATGATCGCGAGGTGACACGAGCGCGCGCTGTGTGGC	2503
Qy	1403	CGAGATCGGAGCCACGGGA---CCTGGCTCGATCCATTCCGATCTCCACGAACGCCGCCAA	1459
Db	2504	CGAGTCCCGGTGCTGAGCCACCCCGGCTCCGTGGATTTCGATTCCTCATCCGCGGCGCG	2563
Qy	1460	TCAGGATGTGCTCTCGCTTGGGACCATGCGCGCGCGCTCTGCGCGCGAGAAGATTCGACCG	1519
Db	2564	AGAGGACCATGCTGTTCATATGGCATGACGGCGGCGCTCAAGGGCGCTCAGGTACAGCACTT	2623
Qy	1520	TTGGGCGGAGATCTTCGCGATCTCTCGCTCTCTGTCTTTGCAACAAGTCCGAGCTCGCGCTG	1579
Db	2624	CGCCGCTTCGTGCTCGCGATTGAATTCCTGTTGGCGCGCAGGCCCTGAGCTTCGCGCT	2683
Qy	1580	CGGACGCGGCTTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGTGTGACAGGCCCTGCGCGA	1639
Db	2684	GCGCTGAAGCCCGGGAAGGCGCCCTCGCGGGGTACGAGCTGG-----TGCGCTC	2734
Qy	1640	GCAGTTCCCGCGCTTGAGACGGACCGGCCCTTGGGACAGGAATTCGCCGCTTGTCTAC	1699
Db	2735	GAAGTCCCCACATGACAAAGNACCGCGAGCTGCACGGGACATCGAGGGGTGAGCCA	2794
Qy	1700	GCACCTC	1706
Db	2795	GCTCGTC	2801

RESULT 3

US-09-252-991A-16475

US-VS-232-991A-16473
: Sequence 16475. Application US/09252991A

; Patent No. 6551795

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES FOR DIAGNOSIS OF AERUGINOSA
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSIS OF AERUGINOSA
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16475
; LENGTH: 1548
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16475

```

	Query Match	9.3%; Score 171.8; DB 3; Length 1548;
	Best Local Similarity 49.2%; Pred. No. 3.6e-25;	
	Matches 606; Conservative 0; Mismatches 577; Indels 48; Gaps 4;	
Qy	237 GGCGGCGCGGGATGTCTTTGCCCTCCGGCGCGACCGTGCCTGCGTCGCGAAGCG 296	
Dd	157 GCCCGGTGGCCTGAGCCTTGATCCAGCGCGACCGCTGATCGCGCCAGCGTCGCC 216	
Qy	297 CGGCTCGGCGCTGTCACTCCGAGGCGCGCATGTCTTACGACTGAACAACCGGCTTCGGT 356	
Dd	217 TGTGTGAGAATCATATCGCGGAGGCGCACCGCTACGGCATCAACACCGGCTTCGCG 276	
Qy	357 CCCCTTGGAAACCGCCTGATCTCAGGTGAGATGTCCGAACGCTGCHAGGCCAATCTTGTC 416	
Dd	277 CTGCTGGCTCGACCGGGATTTCGCCGCGCACTTGGAGAAGTCCACGCTTCATCGTC 336	
Qy	417 CATCATCTGCCACGCGCGTGGGAACCGGTGCTTGACTTGACGACGCGCGCGCCATGTT 476	
Dd	337 CTGTCCCATGCGCGGCTTCGGAGGCGCTTGAACGACGCTATGTTGGCTGCTCATG 396	
Qy	477 CTGGCGGCTCTGTTGTGATCGCTCAAGGAGCCTTCGGTGCAGCGAGGGACCATCGCT 536	
Dd	397 CTGCTCAAGGTGAACAGCCTTGGCGCGCGGTTTCTCCGGCATTCGCCCAAAGTGATCGAC 456	
Qy	537 CGCTCGATCGACCTGTCAATTCCGAGCTCGCTCCGSCCGTTCACAGCCGCGGACCGGTG 596	
Dd	457 GCCTGTATCGCTGTATCAACCGCGAGTCTATCCGCACATCCCGCTGAAGGCTCGGTT 516	
Qy	597 GGCGCGTCGGGTGACCTTGACACCGCTTGGCGATATGTTGCTCTGCTCCAGGGCGGGGA 656	
Dd	517 GGTGCTCCGGACCTGGCGCGCTGGCGCACATGTCTGCT---GGTGTGATCGCGCAA 573	
Qy	657 GATTTCTTGACCGGGAACCGGGAACGGGCTTGAACGGGCTTCGGCGCGGACGG 716	
Dd	574 AGCGGCGCGCCATCGCGTGAATGCTGCCGCGCGCCGAAGCGCTTGGCGGTGGCGCGG 633	
Qy	717 CTCGAACCGCTCGATCTCTCCATCGGATGCACTGSGCGTGTCAACGGGACCTTCGCC 776	
Dd	634 CTGGACCGCTGACCCTTGGCGGCGAAGGAAGGCTTGCTGCTCTCAATGGACCCAGGTG 693	
Qy	777 ATGACCGGATCGCGCTGTGTAATGTCTACGCGCTGCCGCCATCTCGGCAACTGGGCGGTG 836	
Dd	694 TCCACCGCTACGCTTGGCGGGTTTTCGAGGCGGAGGACCTGTTGCGCGCGCCACC 753	
Qy	837 GCGTTGAGCGCCTGTCTTGGGAATGTCTGAGAGCGCGGACCGAGGATGGGCGCGCGCA 896	
Dd	754 GTCTGGCGGGGCTCAGCGTCTGAGGGCCATGCTCGGTTTCGGGGCGCGTTCGATGCGCGC 813	
Qy	897 CTGTCCGACCTGGCGCGCATCCCGACAGAAGACGCCCGCAGCGAGGCTGCGGCGCGCG 956	
Dd	814 ATCCATGCCGCGCGCGCCAGCTGGGCAGATCGACGTGCGCGCGGCTATCGGCACTG 873	
Qy	957 GTGGAACGGCAGCGCGGGTGTTCGGCACGTCAATTGCCGAGCGGAGGGCTGACGCCGCG 1016	
Dd	874 CTTACCGCGCAGCAGGAGGTGGCGCG----- 900	

QY 1017 GATATCGGACGAGCGCGAGCGGGCGAGTGCCTACAGCCTGCGCTGCGCTCCGCGAG 1076
Db 901 ---TCCNATGAGAAGTGGCAGAGGTCCAGAACCCCTATTTCGTGCTGGTTGCCAGCGCGAG 957
QY 1077 GTTCTCGGGCGGGCTTTCGACACGCTCGCATGACCGGTGCTGACGATCGAGCTG 1136
Db 958 GTGATGGGCGCTGCTGACCCAGATGGCGCAGCGCGCGAGGTGCTGGAGATCGAAGCC 1017
QY 1137 AACGCGGTGACGACCAATCCGCTGTTTCCGCCGATGGCAGCGTGCCTGCAAGGG 1196
Db 1018 AACGCGGTGTCGACAAACCGCTGGTATTTCGCCCGCAGGGCG---ACGTGATCTCCGCG 1074
QY 1197 GGCAAATTTCAATGGCGCAGACATGTGGCGCTGACCTCCGATCGCTCGCCACGCGCGCTCACC 1256
Db 1075 GGCAACTTCCACGCCGACCGGTGGCGATGGCGCGCGCAACCTTGGCCCTGGGGTTGGCC 1134
QY 1257 GTTCTCGGGCGCTTTCGCGAGCGCAGATGTCACGCTGACGATGAAGAGGTGAACCGT 1316
Db 1135 GAGATCGGTTCGCTGCGGAAACCGCGCATCTCGCTGATGATGA-----CATGCACATG 1188
QY 1317 GGGCTGCCCCCTTCTCCACCGGGCCCCCGCGGTTGNAATTCGGCTTCATGGGGCGCA 1376
Db 1189 TCGCATGTGCGCGCTTCTTGGTGGCCAAACGCGGGGTCAACTCCGCTTCATGATCGCC 1248
QY 1377 CAGGTGACGGCGACCGCGCTTCTGGCGAGATGCGAGCGCACAGGCGCTGCTCGATCCAT 1436
Db 1249 CAGGTACCGCGCGCGCTTGGCGAGGACACAAAGCGCTTGGCCCATCGGCGCAGGTC 1308
QY 1437 TCGATTCACAGAACCGCGCAATCAGGATG 1467
Db 1309 GACAGCTGCGGACCTCGGCCAACACGGAAG 1339

RESULT 4

US-09-252-991A-16126/c
; Sequence 16126, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMON/ S
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16126
; LENGTH: 2295
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16126

Query Match 9.3%; Score 171.8; DB 3; Length 2295;
Best Local Similarity 49.2%; Pred.No. 3.8e-25;
Matches 606; Conservative 0; Mismatches 577; Indels 48; Gaps 4;
QY 237 GCGCGCGCGCGGATGTCTTGCCTCCGCGCGGACCGGTGCGCTCCGAAGCG 296
Db 1362 GCCCGCTGCGCTGAGCCTGGATCCAGCGCGCGATGATCGCGCGCAGCGTCGCC 1303
QY 297 CGGCTCGGCGCTGTATTCGCGAGGCGCGCATGTCTACGGAATGACAAACCGGCTTCGTT 356
Db 1302 TGTGTGCGAACAATCATCGCGGGCGCGCACCGCTTACGGCATCAACACCGGCTTCGCG 1243
QY 357 CCCCTTTCGAAACCGCTGATCTCAGGTGAGATGTCGAAACGCTGCGAGGCCAATCTTGTC 416
Db 1242 CTGCTGGCTCGAGCGGATTTCCCGCGCGAGCTTGGAGAGCTCCAGGTTCCATCGTC 1183
QY 417 CATCATCTGCGCCAGCGCGGTGGGACCGGTGCTTGAATGGACGACGCGCGCGCCATGGTT 476

Db 1182 TGTCTCCATGCGCGCGCTGCGCGAGGCGCTGGACGACCCATGCTGCGCTGGTCATG 1123
QY 477 CTGGCGGCTGTGGTGTGATCGCTCAGGGAGCTTCGGGTGCGCAGCGAGGGACCATCGCT 536
Db 1122 CTGCTCAAGGTGAACAGCTGCGCGCGGGTTCTCCGGCATCCGCGCAAGGTGATCGAC 1063
QY 537 CGCTGATCGACTGTCTCAATTCGAGCTCGCTCCGCGCGTTCCAGCGCGCGCACGGTG 596
Db 1062 GCGCTGATCGGCTGATCAACGCGGAGGTCTATCCGCACATCCCGCTGAAAGGCTCGGTG 1003
QY 597 GCGCGCTCGGGTGACCTGACACGCTTTCGCTATATGCTGCTCTGCTCCAGGCGCGGGGA 656
Db 1002 GGTGCGCTCCGGGACCTGCGCGCTGCGCGCACATGCTGCT---GGTGTGATCGCGCA 946
QY 657 GACTTCTGACCGGACCGGACCGCGCTTGACGGCGCAGAGGGCTCCGCGCGGACCG 716
Db 945 AGCGGGCGCGCATCGCGGTGAGTGGCTGCGCGCGCGCAAGCGCTGGCGGTGGCGCGG 886
QY 717 CTGCAACCGCTCGATCTCTCCATCGCATGCACTGGCGCTGGTCAACCGGACCTTCGCC 776
Db 885 CTGGAGCGCTGACCTTGGCGGAGGAGGCTTGGCCCTGCTCAATGACACCCAGGTG 826
QY 777 ATGACCGGATCGCGCTGTAATGCTACGCTGCGGCCATCTCGGCAACTGGCGGTG 836
Db 825 TCCACCGCTACGCTGCGCGGTGTTTCGAGGCGGAGGACCTGTTGCGCGCGCGCAC 766
QY 837 GCGTTGACGGCTTCTTGGGNAATGCTCTGAGAGGCGCGACCGAGCATGGCGCGGCA 896
Db 765 GTCTGCGGGGCTCAGCGTCAGGCGCATGCTCGGTTTCGCGGCGCGCTTCGATGCGCG 706
QY 897 CTGTCCGACTGCGCGCGCATCCCGGACAGAGGACGCGCAGCAGAGGTGCGCGCGCGC 956
Db 705 ATCCATGCGCGCGCGCGCAGCGTGGGCAAGTCACTGCGCGCGCTATCGGACCTG 646
QY 957 GTGACCGCAGCGCGGTGCTCGGACGCTCAATTCGCGAGCGGAGGTTCGACGCGCGC 1016
Db 645 CTACCGCGCAGCAGGCTGCGCGC----- 619
QY 1017 GATATCGGACGCGCGGCGGCGAGGATGCTTACAGCTGCGCTGCGCTCCGCGAG 1076
Db 618 ---TCCATGAGAAGTGCACAAAGGTCCAGGACCCCTATTTCGTGCTTGCACGCGCGAG 562
QY 1077 GTTCTCGGGCGGGCTTCGACACGCTCGCATGGCATGACCGGGTGTGACGATCGAGCTG 1136
Db 561 GTGATGGCGCTTCTGCTGACCCAGATGCGCGAGGCGCGGAGGTGCTGGAGATCGAAGCC 502
QY 1137 AACGCGGTGACGACAAATCCGCTGTTTCCGCGGATGGCAGCGTGCCTGCGCGGG 1196
Db 501 AACGCGGTGTCGACAAACCGCTGTTTTCGCGCGCGAGGGCG---ACGTGATCTCCGCG 445
QY 1197 GGCAAATTTCAATGGCGCAGCATGTGGCGCTGAGCTCGATGCGCTCGCCACGCGCGTCA 1256
Db 444 GGCAACTTCCACGCCGAAACCGGTGGCGATGGCGCGCACAACTTGGCTTGGGCTTGGCC 385
QY 1257 GTTCTGCGGGCGCTTCGCGAGCGCCAGATTGACGCTCTGACAGATGAAGGCTGAACCGT 1316
Db 384 GAGATCGGTTTCGCTGCGGAAACCGCGCATCTCGCTGATGATGA-----CATGCACATG 331
QY 1317 GGGCTGCGGCTTCTTCCACCGGGGCGCGCGGGTTGAAATTCGGCTTTCATGGGGCA 1376
Db 330 TCGCAGTTTCGCGCGCTTCTTGGTGGCAACGCGGGGTCAACTCCGCGCTTCATGATCGCC 271
QY 1377 CAGGTGACGCGGACCGCGCTTCCGCGGAGATGCGAGCGCACCGGACCTGCTCGATCCAT 1436
Db 270 CAGGTACCGCGCGCGCTTGGCGAGGACAAACAGGCGCTTGGCCCATCGGCCAGGTC 211
QY 1437 TCGATCTCCACGAACCGCGCAATCAGGATG 1467
Db 210 GACAGCTGCGGACCTCGGCCAACACGGAAG 180

QY 213 GACAGGCCCATGCGTGGCGAGCGCGCGCGCGGATGTCCTTGGCCCTCCGCGCGC 272
DB 49 GACAGCTTATCGCGTTGCGCGCCAGAAAGCCGCATCAGCATTTCTCCGCAAGTACTT 108
QY 273 GACCGTGCCTGCGTCCGAGCGCGCTCGCGCTGTCTATCCGCGAGCGCGCATGTC 332
DB 109 GAGAACTGGCTTCGCTCCGAGCACATATCGATGCACTAGCATCCGCTGATACCCCGGTT 168
QY 333 TAGCGACTGACAAACCGCTTCGTTCCCTTTCGAAACCGCTGTCTCAGGTGAGAAATGTC 392
DB 169 TATGGCAATTCACCGGCTTTGGCGCTTGGCAACCGCCACATCGCAACCGAGATCGC 228
QY 393 CGAAGCTGACGGCCAACTTTGTCATCATCTGCGACGCGCTGGGACCGGTGCTTGAC 452
DB 229 GCCAAGCTGACGCGCTCCCTCATCCGTTCCACGCTGCTGGCATGGGTGAACCGGTGAG 288
QY 453 TGGACGAGCGCGCGCCATGTTCTGGCGGCTGTGGTGTGATCGCTCAGGAGCGCTCC 512
DB 289 CGGAAAGTGTCCGCGCATTTGATGTTCTTGGTGCAGAACCCCTGGCTTCGCGCGCACG 348
QY 513 GGTGCCAGCGAGGACCATCGCTCGCTCATCGACCTGTCTCAATTCGAGCTCGCTCCG 572
DB 349 GCGTTCGCGCGGTTGCTTGAGACCATGTCGBCATGCTCAATGTCAGGCACTCACTCCG 408
QY 573 GCCGTTCCAGCGCGCGACGCTGGCGCTCGGTGACCTGACACCGCTTGGGCAATG 632
DB 409 GTAGTCGCGAATAGGCTTCACTGGGCTGCTCGGTGACTTGGCTCGCTGCTCGACTGC 468
QY 633 GTGCTGTGCTTCAGGCGCGGAGACTTCTCTGACCGGGACGCGCGCTTGAAGGC 692
DB 469 GCATTAAGTGTGATGGCGAGGCGAAGCACCGATGCCACGCGCACATCCCGCGGTA 528
QY 693 GCAGAGGGCTCCGCGCGGACGCTGCAACCGCTGTGATCTCTCCATCGCATGCACTG 752
DB 529 CCGAACTGTTCCGCGGCGGATGATGACCCCTGTGAACTGGCAGAAAGGAGCGCTG 588
QY 753 GCGCTGTGTAACGGGACCTCCGCCATGACCGGATCGCGTGTGTAATGCTCACGCTGC 812
DB 589 GCTCTGTGTAACGGCACCGACGCGATGCTCGGCCAGCTGATCATGCAATTTGGCGGACCTC 648
QY 813 CGCCATCTCGGCAACTGGGCGGTGGGTGAACGCGCTGCTTCCGGAATGTCGAGAGGC 872
DB 649 GATGAGCTGCTGACATCGCCGATGCCATGCCACGCGCCATGAGCGTTGAAGCCCACTGGGC 708
QY 873 CGGACCGAGGATGGCGCGGCACTGTCCGA---CCTGCGCGCATCCCGACAGAG 929
DB 709 ACCGATCAGGATATCCGCGGAGAACTGCAAGAACCACTGGCGCCCGCACCGAGCCAGGC 768
QY 930 GACCGCGACGAGGCTGCGCGCCGCGTGGACGCGCGGCTGGTTCGCGCACGTC 989
DB 769 CGCAGCGCCAGAAATGTTGCGCTTCTGGCCGACTCGCCAAATGTTGCCTCGC---- 823
QY 990 ATTGCGAGCGGAGGCTCGAGCGCGGATATCGGACGAGGCGGAGGCGGCGAGAT 1049
DB 824 -----ATCGGAGGAGAGCGCGC---ATGTCAGGAT 852
QY 1050 GCCTACAGCTGCGTGGCTCGCAGGTTCTCGGGCGGGCTTCGACAGCTCGCATGG 1109
DB 853 GCCTACTCGTGGTTGCTCGCGGAGGTACCGCGCGCGCGCGCACCAATTTGCTCAT 912
QY 1110 CATGACCGGCTGTCGATGCTGAAACCGCGGTGACCGAATCCGGTGTTCGCGCC 1169
DB 913 GCCCGCTTGTTCGCAACCGGAACTGGCTGCGGCATTGACAAACCTGTGGTGTGCCCC 972
QY 1170 GATGGAGGCTGCCCGCTTGCACCGGGGCAATTTATGGGCGAGATGTCGCTGAGC 1229
DB 973 AGCGGGAAGTACTTC-----CAACGGCAACTTCCACGCGGACCGGTAGCTACGTG 1026
QY 1230 TCGGATGCGCTCGCACGCGCTCACCGTTCTGGCGGCGCTTCGAGAGCGCGCATGCA 1289
DB 1027 CTGGAATCTCTTGCCATCGCGTGGCGGACCTCGGCTCTATCGCCGAGCGCGCACCGAC 1086
QY 1290 CGTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGGGCGCGCC 1349

DB 1087 CGCATGCTCGACCCAGCCGCTCCCGGACCTCGCGCATTTCTGGCGGACGATCCGGGT 1146
QY 1350 GGGTTGAATTCGGCTTCATGGCGGCACAGGTGACGGGACCGCGTCTCTGGCGGAGATG 1409
DB 1147 GTGAGCTCGGGCATGATGATCGCCAGTACACCCAGGCGGGCTTGGTGGCAGAAACAAG 1206
QY 1410 CGAGCCACGGGACCTCGCTGATCCATTCGATCTCCACGAACCGCGCAATCAGGATGTG 1469
DB 1207 CGGCTGGCAGTTCCTGCGCAGCGTTGACTCCATCCCATCTCGGCCCATGAGGAGACCAC 1266
QY 1470 GTCTCGCTTGGGACCATCGCGCGCGC 1496
DB 1267 GTTTCCTGGCTGGCATCGCGCGCGC 1293

RESULT 7

US-09-833-745-12
; Sequence 12, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic expression
; OTHER INFORMATION: vector sequence
US-09-833-745-12

Query Match 8.3%; Score 153.8; DB 3; Length 1536;
Best Local Similarity 48.9%; Pred. No. 1.4e-21;
Matches 629; Conservative 0; Mismatches 607; Indels 51; Gaps 6;
QY 213 GACAGGCCCATGCGCTGGCGAGCGCGCGCGGATGTCCTTGGCCCTCCGCGCGC 272
DB 49 GACGAGCTTATCGCGCTTGGCCGCGACGAAGCCGCATCAGCATTTCTCCGCAAGTACTT 108
QY 273 GACCGTGCCTGCGTCCGAGCGGCTCGGGCTGTCTATCGCGAGGCGCGCATGTC 332
DB 109 GAGGAATGCTGCTCCGTCGCGACCATATCGATGCACTAGCATCCGCTGATACCCCGGTT 168
QY 333 TAGCGACTGACAAACCGCTTCGCTCCCTTGGAAACCGCTGTCTCAGGTGAGAAATGTC 392
DB 169 TATGGCAATTCACCGGCTTTGGCGGTTGGCAACCCGCAATCGCACCGGAGATCGC 228
QY 393 CGAAGCTGACGGCCAACTTGTTCATCATCTGGCGAGCGGCTGGGACCGGTGCTTGAC 452
DB 229 GCCAAGCTGACGCGCTCCCTCATCCGTTCCACGCTGTCTGGCATGGGTGAACCGGTGAG 288
QY 453 TGGACGAGCGCGCGCCATGTTCTGGCGGCTGTGGTGTGATCGATCGCTCAGGAGCGCTCC 512
DB 289 CGGAAAGTGGTCCGCGCATTTGATGTTCTTGGCTGCAAGACCCCTGGCTTCGCG---CCGC 345
QY 513 GGTGCCAGCGAGGGAACCATCGCTCGCTGATCGACTGTCTCAATTCGAGCTCGCTCCG 572
DB 346 AGGTTCCGCGGTTGTCTTGGAGACCATGGTGGCATGCTCAATGCGAGCATCACTCCG 405
QY 573 GCGCTTCCAGCGCGGACGCTGGGCGGCTCGGTGAGCTGACACCGCTTGGCAGATG 632
DB 406 GTAGTCCGGAATAGGTTCACTGGGCTGCTCCGGTGACTTGGCTCCGCTCGGCACTGC 465

Qy	633	GTGCTCTGCTCCAGGCGCGGGGAGACTTCTGGACCGGGACGGGACCGCGCTTGTACGGC	692
Db	466	GCAATTAGTGTGATGGGCGAGGGGGAAGCCACCGATGCCACGGCGACATCCGCCCGGTA	525
Qy	693	GCAGAAGGGCTCCGCGCGGACGGCTGCAAACGGCTCGCATCTCTCCATCGCGATGACATCG	752
Db	526	CCGGAACTGTTTCGCCGAGCGCGATTGACCCCTGTGGAATGGGCAGAAAAGGAGGCCCTG	585
Qy	753	GGCTGTGTTCAACGGGACCTCCGCCAATGACCGGGATCGCGCTGGTGGAAATGCTCAACGCCTGC	812
Db	586	GCTCTGTGTCAACGGCACCGACCGGATGCTCGCGCCAGCTGATCATGGCATTTGGCGGACCTC	645
Qy	813	CGCCATCTCGGCAACTGGGCGGTGGCGTTGAGCGCCCTGCTTGGGATGTCGAGAGGC	872
Db	646	GATGAGCTGTGGACATCGCCGATGCCACCGCGCCATGACGCTTGAAGCCAGCTGGCG	705
Qy	873	CGGACCGAGGATGGGCGCGGCACTGTCCGA----CTGCGCGCGCATCTCCGGACAGAA	929
Db	706	ACCGATCAGGTATTTCGGCGCAGAACTGCACGAAACAATGCGCCCGCACCCAGGCGCAGGC	765
Qy	930	GACGCGGACGAGGCTGCGCGCCGCGTGTGACGCGCAGCGCGCGGTGGTTCGGGACGTC	989
Db	766	CGCAGCGCCCAAGAACATGTTCCGCTTCTTGCGCGACTCGCCAAATTGTTGCCCTCGC-----	820
Qy	990	ATTGCCGAGCGGAGGCTCGACGCGGCGGATATCGGGACGGAGCGGAGGCGGGGACGAT	1049
Db	821	-----ATCGGAGGGAGACGGCGG--AGTGCAGGAT	849
Qy	1050	GCCTACAGCTCGCTGCGCTCCGAGGTTCTCGGGCGGGCTTCGACACGCTCGCATGG	1109
Db	850	GCCTACTGCTGCGTTGCTCGCCGAGGTACACCGCGCGCCCGGACACCATTTGCTCAT	909
Qy	1110	CATGACCGGGTGTGACGATCGAGCTGAAACGCGGTGACCGACAAATCGGTGTTTCGCGCC	1169
Db	910	GCCCGCTGTGTGCGCACCGCGAACTGCGTCGCGCCATTGACAAACCTGTGGTGTGCGCC	969
Qy	1170	GATGGCAGCTGCCGCGCTGCAAGGGGGCAATTTATGGGCGCAGCATGTGGCGCTGACG	1229
Db	970	AGCGCGAAGTGACTTC-----CAACGGCAACTTTCACGGCGCACCGGTAGCTACGTG	1023
Qy	1230	TCCGATGCGCTCGCACCGGCGGTCAACGCTTCTGGCGGGCCCTTGGGAGCGCCAGATTGCA	1289
Db	1024	CTGGACTTCCTTGCCATCGCGTGGCGGACCTCGGCTCTATCGCGAGCGCGCACCGAC	1083
Qy	1290	CGTCTGACAGATGAAGGCTGAAACGGTGGGTGCCCCCTTCCTTCACACGGGGGCCCGCC	1349
Db	1084	CGCATGCTCGACCCAGACCGCTCCCGGACCTGACCGGCAATTCCTGGCGCAGCATCGGGT	1143
Qy	1350	GGGTTGAATTCGGCTTCATGGGCGCACAGGTGAGCGGCGACCGCGCTCTGGCCGAGATG	1409
Db	1144	G---TGAGATCGGGCATGATGATCGCCGAGTACACTCAGGCCCGCTTGGTGGCAGAAAC	1200
Qy	1410	CGAGCCACGGACCTTGCTCGATCCATTGATCTTCACGACGCGCCCAATCAGGATGTG	1469
Db	1201	AAGCGGCTGGCAGTTCCTGAGTTGATCTCCATCCATCTCTCGGCGCATGCGGAAGACAC	1260
Qy	1470	GTCCTGCTGGGACCATGCGCGCGCGC	1496
Db	1261	GTTTCCCTGGCTGGCATGCGGCGCGC	1287

RESULT 8

RESULT 8
 US-09-252-991A-16124/c
 ; Sequence 16124, Application US/09252991A
 ; Patent NO. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18

Db 720 TCAGATACCGCAGATCCATGCGCGCTGCCGTGACCAAGTTGGCCCAACGCGGCGCGCAGATA 661
Qy 1125 ACGATCAGCTGAACGGGTGACCGACAATCCGCTGTTTCGCG---CCGATGGCAGCGTG 1181
Db 660 GAGACCGAATCAACTCGGCCACCGACAACCCGCTGCTGCTGGGTACCGCGAAGCCTAC 601
Qy 1182 CCGCCCTGCACGGGGCAATTTTCATGGGCGCAGCATGTGGCGCTGAGCGTCCGATCGCGCTC 1241
Db 600 CGGTGTGATCCCAAGGCAATCCCAACCGGCAATCCGCTGGCGATGGCGCGAGCTGCTG 541
Qy 1242 GCCAGGCGCTCACCGTTCTGGCGGGCTTTCGAGAGCGCCAGATTGACAGCTGTGACAGAT 1301
Db 540 GCGATCCCGTGGCGAGCTGGCGGGGTGCGCGAGCGGGCTCGATCGCCTG---GTC 484
Qy 1302 GAAAGGCTGAACCGTGGGCTGCCCCCTTCTCTCCACCGGGGCCCGCGGCTTGAATTC 1361
Db 483 AACCGCTGTGACGGCTTGCAGGCGTTCCT---GGTGGCAAGCCGGAGTCAACTCG 427
Qy 1362 GGCTTCATGGGCGCACAGGTGAACGGGACCGCGCTCCTGCGCGAGATGCG---AGCCACG 1418
Db 426 GGGATGATGATCACCAAGTAGCTGCGCGCTCCCTGCTGGCGAGAACCGCCAGCTGGCG 367
Qy 1419 GGACCTGCTCGATCAATTCGATTCACAGAACGCCGCCAATCAGGATGTGCTCGCTT 1478
Db 366 CAGCCGCGGTGTCGACAACTTCGTCACTCCGCGCTCCAGGAGGACCACTGAGCCTC 307
Qy 1479 GGAACCATCGCGCGCTCTGCGCGAGAGATGCAACCGTTGGGCGAGATCCTTGCG 1538
Db 306 GGCACAGTCCCGGCTCAAGCTTGGCAGGCGCTCGAATTCCTGCCCGCAACGCTTGGCCAG 247
Qy 1539 ATCCTGCTCTCTGTGTCACAACTGCGGAGTGGCGCTGCGCGACGCGCTTAGACGGG 1598
Db 246 ATCAGTACCTGTGGCGCGCCAGGCTTCGAATTCCTGCCCGCAACGCTTGGCCAG 187
Qy 1599 GTGTCTCCCGGGGAAGAGCTGTGTCAGCGCCCTGCGCAGCAGTTCGCGCGCTTGAG 1658
Db 186 GGCACCGCGCGCTGGGCAT-----CCTGCGCGAGCGGTGCGCGGTACGAC 136
Qy 1659 ACGACCGCGCCCTGGGACAGGAATTCGCGCTTGTACGCACTCTTGGACG 1713
Db 135 ACCGACCGTGGCTGCTCCGACATCGCAGCGCGCGGCCCATCTCTCGCGCAGC 81

RESULT 9
US-09-252-991A-16478
; Sequence 16478, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16478
; LENGTH: 1563
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16478

Query Match 6.8%; Score 125.4; DB 3; Length 1563;
Best Local Similarity 47.8%; Pred. No. 5.9e-16;
Matches 714; Conservative 0; Mismatches 706; Indels 75; Gaps 9;

Qy 225 GCCGTGGCAGCGCGCGCGGATGTCCTTGGCCCTCCGCGGCGGACCGGTCGCT 284
Db 91 GCGGTGGCCCGCAGCGGTGCGCGCTGGAATCTGTCGCGCGCGCCTGGGCAACGATCGAC 150

Qy 285 GCCTCGAAGCGCGCTCGCGCTGTATCATCGCGAGCGCGCCATGTCTACGACGTGACA 344
Db 151 AACGCCCGGCCATCGTTCGCCGTATCGTCGCAACGCGGAGCGTGCCTACGGCATCAGT 210
Qy 345 ACGCGCTTGGTCCCTTTCGGAACCGCCTGATCTCAGGTGAGAAATGTCGGAACGCTGCGAG 404
Db 211 ACCGCGCTCGCGCGCTGTGCGAGCTGCTCTGGAAGGCGAGCAGCTCGCCGAGCTGTGCG 270
Qy 405 GCCAATCTTGTCCATCATCTGGGCGAGCGGCGTGGGACCGGTGCTTACTGGACGAGGCG 464
Db 271 CGCAACACCTTGTCTACGCCATGCTTCGGGGTTCGGGAGCCCTCGCGGACGAAACAGACC 330
Qy 465 CGCGCATGCTTCTGCGCGGTCTGTGTGATCGCTCAGGGAGCCTCCGCTGCGCAGCGAG 524
Db 331 CGGCGCATCATCTGTGCGCGGTTCGCAACTACAGCAGGGGCAAGTTCGGGCTCGACCGT 390
Qy 525 GGAACCATCTGCTCGCTGATTCGACCTGCTCAATTCGAGCTCGCTCGGCGGCTTCCGAGC 584
Db 391 TCGCTGGTGAAGGACTGCTGCGCTGCTCAACACCGCATTAACCCCGCAGGTGCGCGGCC 450
Qy 585 CGCGCACCGTGGCGCGCTCGGTGACCTGACACCGCTTGGCGATATGCTGCTCTGCTC 644
Db 451 CAGGGCTCGGTGGG-----TACCTGACCCACATGGCGCAGCTCGGCATCGCCCTG 501
Qy 645 CAGGGCGGGGAGACTTCTTGGAACCGGAGCGGCTTTCGCGCGCAGAGGGGCTC 704
Db 502 CTCGGCATCGCGAGGTGAGCTACCGCGGAGCGCTGTCGCGCGCGCGC---CGGTTG 558
Qy 705 CGCGCGGACGGCTGCAACCGCTGATCTCTCCATTCGCGATGCACTGGCGCTGGTCAAC 764
Db 559 GCGCGGAAGGCGCTGCGAGCGGTGCGCTTGGGAGCCAGGACGCGGCTCTGCTGTGTCAAC 618
Qy 765 GGGACCTCCGCCATGACCGGATCGCGCTGGTGTGATGCTCAGCGCTCGCGCCATCTCGGC 824
Db 619 GGCACCGCTGATGACCGGCTCGCTCGCTGCGCGCTTGGACGACGCGCGCTGGCG 678
Qy 825 AACTGGCGGTGCGCTGACGCGCTCTTGGGAAATGCTGAGAGCGCGGACCGGAGGCA 884
Db 679 CAGTGGCGCGAGTGTGATCGGGCGATGAGCTTCGAGGCGCTTGGCGGCGCAACTGGCGGCG 738
Qy 885 TGGGCGCGGCACTGTTCGACCTGCGCGCGCATTCGCGGACAGAAAGACGCGCGCAGGAGG 944
Db 739 TTGACGCGCGAGATCTTGGCTCAAGCGCATCCCGGATGCGAGCGGTGCGCGCTCAAT 798
Qy 945 CTGCGCGCGCGTGGACGCGGCGCGGTGGTCCGSCACGTCTATTGCCGAGCGGAGG 1004
Db 799 CTGCGAGCCTTGTGCTGGCAGCCAGGTACTGGAGA----- 835
Qy 1005 CTCGACGCGCGGATATCGGAGCGGAGCGGCGGCGGCGGAGGATGCTTACAGCCTGCGC 1064
Db 836 ----ACGCGCGGCGCATCCGAC-----CCAGGATGCCCTGAGCATCCGC 876
Qy 1065 TGGCTTCGCGAGTTCCTCGGGCGGCGCTTCGACACGCTCGCATGGCATGACCGGCTGCTG 1124
Db 877 TCGATACCGCAGATCATGCGCGCTGCGCTGACCAAGTTGGCGCCACGCGCGCGCGCAGATA 936
Qy 1125 ACGATCGAGCTGAACCGGCTGACCGGCAATCCGCTGTTTCCGCG---CCGATGSCAGCTG 1181
Db 937 GAGACCGAACTCAACTCGGCCACCGGACACCGCGTCTGCTGGGTACGCGCGGAGCCTAC 996
Qy 1182 CCGCGCCTGCACGGGGCGAATTTTCATGGGCGCAGCATGTGGCGCTGACGCTCCGATGCGCTC 1241
Db 997 CGGTGTGATCCAGGCGCAATCCCGACGCGGAATCGTGGCGGATGGCGCGGACCTGCTG 1056
Qy 1242 GCACGGCGCTCACCGTTCTGGCGGCGCTTTCGAGAGCGCCAGATTGCACTGTGACAGAT 1301
Db 1057 GCGATCGCGCTGCGCGAGCTGGCGGGGTGCGCGAGCGGCGCTGATGCGCTG---GTC 1113
Qy 1302 GAAAGCTGAACCGTGGGCTGCCCTTCTCCACCGGGGCGCGCGGTTGAATTC 1361
Db 1114 AACCCGCTGTCAGCGGCTGCGCGGTTCCT---GGTGGGCAAGCCGGAGTCAACTCG 1170
Qy 1362 GGCTTCATGGGCGCAGAGTGAACGCGGACCGCGCTCTTGGCGGAGATGCG---AGCCACG 1418

Db 1171 GGGATGATATCACCAGTACGTGCGCGCTTCCCTTGGCTGCGGAGAACCGCCAGCTGGCG 1230
Qy 1419 GGACTGCTCGATCCATTCGATCTCCACGAAACCGCCCAATCAGGATGTGGTCTCGCTT 1478
Db 1231 CAGCGCGGGTGGTCGACAACTTCGTCACTTCGCGCTCCAGAGGACCACTTACGCTC 1290
Qy 1479 GGGACCATCGCGCGCTCTCGCGGAGAGATCGACGTTGGGCGGAGATCTTGGC 1538
Db 1291 GGCACCAAGTCCGCGCTCAAGCTTGGCAGGCGCTTGGAGAACCTCCGCGGATCTCGCC 1350
Qy 1539 ATCTCTGCTCTCTCTTTCGACAAAGCTTCGCGAGCTTCGCGCTTCGCGAGCGGCTAGACGG 1598
Db 1351 ATCAGTACCTGCTGGCGCCAGCGCTTCGAATTCCTGCGCCCGCAACGCTTCGGCAG 1410
Qy 1599 GTGTCTCCCGCGGAGAAAGCTGGTGCAGGCCCTTCGCGAGCAGTTCGCCCGCTTGAAG 1658
Db 1411 GGCACCGCGCGCTGGCGCAT-----CTGCGCGAGCGCTGCGCGTACGAC 1461
Qy 1659 ACGACCGCGCTGGGACAGGAATTTGCCCGCTTGTACGCACTCTTGCAGC 1713
Db 1462 ACCACCGCTGGTGGCTCCCGACATCGCCAGCGCGCGCATCTCTCGCGAGC 1516

RESULT 10
US-09-624-693A-16
; Sequence 16, Application US/09624693A
; Patent No. 6355468
; GENERAL INFORMATION:
; APPLICANT: Yoshida, Roberta
; APPLICANT: Kootstra, Anna
; TITLE OF INVENTION: Phenylalanine Ammonia Lyase Polypeptide and
; TITLE OF INVENTION: Polynucleotide Sequences and Methods of Obtaining and
; TITLE OF INVENTION: Using Same
; FILE REFERENCE: 29479/500NSC
; CURRENT APPLICATION NUMBER: US/09/624,693A
; CURRENT FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Rhodotorula mucilaginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (646)..(2784)
US-09-624-693A-16

Query Match 5.0%; Score 110.4; DB 3; Length 2787;
Best Local Similarity 46.6%; Pred. No. 5.9e-13;
Matches 514; Conservative 0; Mismatches 566; Indels 24; Gaps 4;

Qy 483 CGTCTGCTGATCGCTCAGGGAGCTTCGCGTCCAGCGAGGGAGCATCGCTCGCGCTG 542
Db 1177 CGTGTCACTCGCTCATCTCGGGTCACTCGCGGTTCGCGATCGCTCTCGAAGCCCTC 1236
Qy 543 ATCGACTGCTCAATTCGAGCTCGCTTCGCGCGGTTCCTCCAGCGCGACAGTGGGCGG 602
Db 1237 ACCAACTTCCTCAACACGAGCATCACCCGATCGTCCCGCTTCGAGGACCACTTCGGCG 1296
Qy 603 TCGGTGACCTGACACCGCTTCGCGATATGTTGCTTCGCTTCACGCGCGGGAGACTTC 662
Db 1297 TCGGGGACCTTTCCCGCTCTTTACATCGCGCTTCGATCAACCGCCACCGGACTCG 1356
Qy 663 CTGG--ACCGGAGCGGAGCGGCTTTGACGGCGCAGAGGCTTCGCGCGGACGCGCTG 719
Db 1357 AGGTCCAGCTCGACGCGAAGATCATGTCCGCCAGGAGGATTCGGCTCAAGGGTCTT 1416
Qy 720 CAACCGCTCGATCTCTCCCATCGCGATGCACTGGCGCTGGTTCACAGGGAACCTTCGCCATG 779
Db 1417 CAGCCCGTCTGCTCGGTCCGAGAGGGTCTCGGTCTCTGATCAACGCGCAGCGCTCC 1476
Qy 780 ACCGGGATCGCGTGGTGAATGTCTACGCTTCGCCCATCTTCGGCAACTTCGGCGGTGGCG 839

Db 1477 GCCTCGATGCGAGCGCTGGCCCTTACCAGCGCACAGTCTCTCTCGCTCTCGCACAGCGC 1536
Qy 840 TTGACGCGCCCTGCTTGGGAATGTCTGAGAGCGCGACCGAGGCGATGGGCGCGGCACTG 899
Db 1537 CTCACTGCTCTTACTGTGAGGGCCATGGTCGACACGCGCGGCTGTTCCACCACTTCTC 1596
Qy 900 TCGACACT---GGGCGCGATTCGCGGACAGAGGACGCGCGAGGAGGCTGCGCGCGCGC 956
Db 1597 CAGGACGTCAAGCGGCTTACCGCGACCCAGATCGAGGTGGCGGCAACATCCGGACTCT 1656
Qy 957 GTGAGCGGAGCGCGCGGTGGTCCGACAGTTCATTCGCGAGCGGAGGCTCGACGCGCGC 1016
Db 1657 CTGAGGCGAGCAAGTACGCGCTCCACCGAGACTGAAGTCAAGGTCAAGGAGCAGAG 1716
Qy 1017 GATATCGGAGCGAGCGCGGAGGCGGAGATGCTTACAGCTTCGCTGCGCTTCGCGCAG 1076
Db 1717 GGCATCTCTCA-----GGCAGGACCGGTACCGCTCGCTGCTCGCGCGAG 1761
Qy 1077 GTTCTCGGCGCGGCTTCGACACGCTCGCATGGAATGACCGGGTGTGAGATCGAGGTG 1136
Db 1762 TGGCTCGGTCTCTGTCAGCGACATGATTCACGCTCACGCTGTCTCTCTCGCTCGAGGT 1821
Qy 1137 AACCGGTGACGACCAATCCGGTGTTCGCGCGATGCGAGCGGTGCGCGCTTCGACGCGG 1196
Db 1822 GGTGAGTCGACCAACCGACAAACCGCTGATCGACTCGAGAACAGATGACCCACCATGGC 1881
Qy 1197 GGCAAATTTTCATGGGCCAGCATGTGGCGCTGACGTCCGATCGCTTCGCGACGCGCGTCA 1256
Db 1882 GGAGCTTTCATGGCGAGCAGCGTCGGAACACAGATGGAGAGACTCGCTCGCGCTCGCG 1941
Qy 1257 GTTCTGGCGGCGCTTCGCGAGCGCCAGATGTGACGTCTGACAGATGAAGGTGAACCGT 1316
Db 1942 CTGATGGCAAGGTCACTTACTAGCTCACCGAGATGCTCAACGCGCGCATGAACCGG 2001
Qy 1317 GGGTGGCGGCTTCCTCCACCGGGCGCGCGGGTGAATTCGGCTTCATGGGCGCA 1376
Db 2002 GCGCTTCGCTGCTGCTGCGTTCGCGAGGACCTTCCTCTTATCATCTCAAGGGTCTC 2061
Qy 1377 CAGGTGACGCGGACCGCGCTCTGCGCGAGATGCGAGCC---ACGGGACCTGCTCGATC 1433
Db 2062 GACATTCGTGCGCGCGCTTACACTTCGAGCTCGTTCACCTTCGCAACCGGTTTCGACC 2121
Qy 1434 CATTCGATCTCCACGAAACCGCGCAATCAGATGTGTCTGCTTGGGACATCGCGCGC 1493
Db 2122 CAGTTCAGCGCGCGCGAGATGGGCAACCGAGCCATCAACTCGCTCGCTCTCATCTCGGCC 2181
Qy 1494 CGCTTCGCGCGAGAGATCGACCGTTGGGCGGAGATCCTTCGATCTCGCTCTCTGT 1553
Db 2182 CGCGGACCGCGCGAGGAGCAACGAGCTTCTCTCCCTCTCTCGCCACCCACCTCTACTGC 2241
Qy 1554 CTGCGCAAGCTCGGAGCTGCGC 1577
Db 2242 GTCTTCAGGCGCTCGACCTCGC 2265

RESULT 11
US-09-252-991A-16368
; Sequence 16368, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16368

QY 1254 ACCGTTCTGGCGGCTTTCGGAGCCGAGATTGACAGTCTGACAGATGAAAGGCTGAAC 1313
DB 1312 GCCCTGATCGGCAAGCTCAACTTACGAGCTACACGAGATGCTCAACGCCGCGATGAAC 1371
QY 1314 CGTGGGCTGCCCTTCTCTCCACCGGGCCCGCGGGTTGAATTCGGGCTTCATGGGC 1373
DB 1372 CCGGCTTNCNTCTGCTCGTCCGAGGACCCNTCGCTCTCTCTATCACTGCAAGGCG 1431
QY 1374 GCACAGGTGACGGCACCGCGCTCTGCGCGAGATGCGAGCC---ACGGAGCTTGCCTCG 1430
DB 1432 CTCGACATTCGCGCGNCGNCTACATTCGAGGCTCGGACCTTTCGCAACCCGCTNACG 1491
QY 1431 ATCCATTGATCTCCACGAACGCCGCCCAATCAGGATGTGCTCTGCTTTGGAGCATGGCC 1490
DB 1492 ACCAAGTCCAGCGCGNGAGATGGGCAACAGGCGCTCAACTCGCTCGNCTCATCTCG 1551
QY 1491 GCGGCTCTGCGCGAGAGATCGACGTTTGGCGGAGATCTTTGCGATCTCGCTCTC 1550
DB 1552 GCGGCGCACGCGGCGGCAACGAGCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1611
QY 1551 TGTCTTGACAGCTGCGGAGCTGCGC 1577
DB 1612 TCGTNTCTCCAGGCGTCACTCGC 1638

RESULT 13

US-09-765-873A-31
; Sequence 31, Application US/09765873A
; Patent No. 6521748
; GENERAL INFORMATION:
; APPLICANT: Tang, Xiao-Song
; TITLE OF INVENTION: BIOPRODUCTION OF PARA-HYDROXYCINNAMIC ACID
; FILE REFERENCE: BC1009 US CIP
; CURRENT APPLICATION NUMBER: US/09/765,873A
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: US 09/627,216
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 60/147,719
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 31
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: mutant from Rhodotorula glutinis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2061)
US-09-765-873A-31

Query Match 5.6%; Score 103.2; DB 3; Length 2061;
Best Local Similarity 46.5%; Pred. No. 1.5e-11;
Matches 511; Conservative 0; Mismatches 553; Indels 36; Gaps 4;

QY 499 CTCAGGAGCTCCGGTCCGACGAGGGGACCATCGCTCGCTCATCGACCTGCTCAATT 558
DB 440 CCCGCGCCACTCGGCTGTCGCTCTGCTGCTCGAGGCGCTCACCACCTTCTCAACC 499
QY 559 CCGAGCTCGCTCCGCGCTTCCACAGCCGCGCACGCTGGGCGCTCGGCTGACCTGACAC 618
DB 500 ACGGATCACCCCATGTCCTCCCTCGCGGACCATCTCTGCTCGGCGACCTGTCTC 559
QY 619 CGCTTGGCGATATGCTGCTCTGCTTCCAGGGCCGGGAGACT-----TCC 663
DB 560 CTCTCTCTTACATTTGCAGCGCCATCAGCGTCTACCCGACAGCAAGTGCACTGCTGC 619
QY 564 TGGACCGGACGAGCGGCTTACCGCGCAGAGAGGCTCCGCGGCGAGCGGTGCAAC 723
DB 620 ACGAGGCAAGGAGAAGATCTGTACGCCCGCGAGGCGATGGCGCTCTTCAACCTCGAGC 679
QY 724 CGCTCGATCTCTCCCATCGCGATCACTTGGGCTGTGTTCAACGAGACCTCCGCGCATGCC 783
DB 680 CCGTCTGCTCTCGGCGCGGAGAGGTTCTGCTCTGTTCAACGCGACCGCGCTCTCAGCAT 739

QY 784 GGATCGCGCTGGTGAATGCTCAAGCTTCCGCTCCGCTTCTCGGCAACTGCGGCGGTGGGTTGA 843
DB 740 CGATGGCCACCTCGCTCTGACGACGCTCATATGCTCTCTCGAGTCGCTCA 799
QY 844 CCGGCTTGTTCGGAATGCTGAGAGCGCGGACCGAGGATGGCGCGGACACTGTCCG 903
DB 800 CCGGATGACGCTGAGAGGATGGTGGGCGAGCGCGGCTGTTCCACCCCTCTTCAAG 859
QY 904 ACCT---GCGGCCGATCTCCGGAAGAGGACGCGGAGGCTGCGCGCCCGCGTGG 960
DB 860 ACGTCAAGCGGCTTTCGCTTCCACCATGAGGAGGAGTCAAGGTCAAGGACGACGAGGGA 979
QY 961 ACGGAGCGCGGCTTGGTCCGCGACGCTATGCGGAGCGAGGCTGACGCGCGGATTA 1020
DB 920 AGGAAAGCGCTTTCGCTTCCACCATGAGGAGGAGTCAAGGTCAAGGACGACGAGGGA 979
1021 TCGGAGCGGAGCGGAGGCGGCGGAGGCTTACGCTCGCTCGCTCGCTCGCTCGCTCGCT 1080
DB 980 TT-----CTCGCGGAGACCGCTACCCCTTGGCGACGCTCTCTCATGTGGC 1024
QY 1081 TCGGCGCGGCTTCGACACGCTCGCATGCGATGACCGGCTGCTGACGATCGAGCTGAACG 1140
DB 1025 TCGGCGGCTCGTCAAGGCTTTCAGCGCCAGCGGCTCTTCCACCATGAGGCGCGCC 1084
QY 1141 CGGTGACGAGCAATCCGCTTTCGCGCGAGGCTGCGCTCGCTCGCTCGCTCGCTCGCT 1200
DB 1085 AGTCGAGCGGCGGACCAACCTCTCATCGACGCTCGGAGAACAGACTTCGACCAACGCGGCA 1144
QY 1201 ATTTATGGGCGGACGATGCGGCTGCGCTCGCATGCGCTCGCGCGGCTCGACCGCTTC 1260
DB 1145 ATTTTCAGGCTCGCTGTCGCGGCTGTCGCGGAGGAGACTCGCTCGGCTCGGCTCGG 1204
QY 1261 TGGCGGCGCTTTCGAGGCGCGGATGCGATGCGATGCGGCTGCGGCTGCGGCTGCGG 1320
DB 1205 TCGGCAAGCTCAACTTCAGCGCTCACCGAGATGCTCAACCGCGGATGAACCGCGGCC 1264
QY 1321 TGCCCCCTTCTTCCACCGGCGGCGGCTTGAATTCGGCTTATGGGCGGACAGG 1380
DB 1265 TCCCTCTCTGCTCGCGGCGGAGACCCCTGCTCTCTTACCACTGGAAGGCGCTCGACA 1324
QY 1381 TGAGGCGGCGGCTCTGCGCGGAGATGCGA---GCCAGGGAGCTCGCTCGATCCATT 1437
DB 1325 TCGCGCTCGGCGGTAACCTCGGAGTTGGGACACCTCGCGAACCTTGTACACGCGATG 1384
QY 1438 CGATCTCCAGGAACCGCGGCAATCAGGATGTGCTCGCTTGGGACCATCGCGCGGCC 1497
DB 1385 TCCAGCGGCTGAGATGCGGACCAAGCGGCTCAACTGCTTGGCTCATCTCGGCTCGTC 1444
QY 1498 TCTGCGCGGAGAGATCGACCGTTGGGCGGAGATTCCTTGGGATCCTCGCTCTCTGTG 1557
DB 1445 GCACGACCGAGTCCACGAGGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1504
QY 1558 CACAAGCTCGGAGCTGCGC 1577
DB 1505 TCCAAGCATCGACTTGGC 1524

RESULT 14

US-09-627-216A-7
; Sequence 7, Application US/09627216A
; Patent No. 6368837
; GENERAL INFORMATION:
; APPLICANT: Sariaelani, Sima F
; APPLICANT: Tang, Xiao-Song
; APPLICANT: Qi, Wei Wei
; APPLICANT: Vannelli, Todd
; APPLICANT: Gatenby, Anthony
; TITLE OF INVENTION: Bioproduction of para-Hydroxycinnamic Acid
; FILE REFERENCE: BC1009 US NA
; CURRENT APPLICATION NUMBER: US/09/627,216A
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/147,719

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 06:02:09 ; Search time 1553 Seconds
(without alignments)
9850.837 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgaggcgccatccacgag.....tcagctcgatcccgacgagg 1850

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134589005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1850	100.0	1850	7	US-10-621-826-2
2	1580	85.4	1581	10	US-11-069-633-7
3	1446	78.2	1446	6	US-10-369-493-31503
4	561	20.3	1626	7	US-10-464-609-3
5	244.2	13.2	1587	6	US-10-369-493-32499
c	242.4	13.1	42000	7	US-10-159-257A-1
c	242.4	13.1	63158	6	US-10-292-198-1
8	238.4	12.9	1497	6	US-10-369-493-31670
9	232.8	12.6	1542	6	US-10-156-761-3312
c	232.8	12.6	9025608	6	US-10-156-761-1
11	231.6	12.5	1521	7	US-10-282-122A-13842
12	215.4	11.6	1503	6	US-10-369-493-42898
13	211.6	11.4	1503	6	US-10-369-493-28443
14	211.6	11.4	1503	6	US-10-369-493-31202
15	211.6	11.4	1521	7	US-10-282-122A-13653
16	177.4	9.6	1518	6	US-10-369-493-28430
17	177.4	9.6	1518	6	US-10-369-493-31190
18	177.2	9.6	1575	6	US-10-369-493-47139
19	173.4	9.4	1530	3	US-09-815-242-7987
20	173.4	9.4	1530	6	US-10-369-493-40527
21	173.4	9.4	1530	7	US-10-282-122A-30600
22	172.8	9.3	2256646	7	US-10-470-565-1
23	170.8	9.2	1530	6	US-10-369-493-35187

24	170.8	9.2	1533	6	US-10-369-493-38162
25	170.8	9.2	1533	6	US-10-369-493-38793
26	170.8	9.2	1542	6	US-10-369-493-36026
27	161.8	8.7	1542	3	US-09-833-745-7
28	161.6	8.7	1521	6	US-10-369-493-39101
29	161.6	8.7	1521	6	US-10-369-493-39469
30	161.6	8.7	1521	6	US-10-369-493-39852
31	159.4	8.6	1566	6	US-10-369-493-32807
32	157	8.5	1525	7	US-10-282-122A-23382
33	155.2	8.4	1545	7	US-10-282-122A-33163
34	153.8	8.3	1536	3	US-09-833-745-12
35	151.8	8.2	1446	6	US-10-369-493-28511
36	151.8	8.2	1491	6	US-10-369-493-31270
37	151.6	8.2	1452	6	US-10-369-493-32328
38	149.4	8.1	1530	6	US-10-369-493-37377
39	146.6	7.9	1530	7	US-10-282-122A-31973
40	144.6	7.8	1461	6	US-10-369-493-27975
41	144.6	7.8	1506	6	US-10-369-493-30731
42	144.6	7.8	1533	7	US-10-282-122A-12902
43	141.2	7.6	1362	6	US-10-369-493-37343
44	140.8	7.6	1386	6	US-10-369-493-42300
45	133.4	7.2	1521	6	US-10-369-493-32286

ALIGNMENTS

RESULT 1

US-10-621-826-2
; Sequence 2, Application US/10621826
; Publication No. US20040059103A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company, Inc.
; APPLICANT: Huang, Lixuan
; APPLICANT: Xue, Zhixiong
; TITLE OF INVENTION: DNA and Amino Acid Sequences of a Tyrosine Ammonia Lyase Enzyme
; TITLE OF INVENTION: From the Bacterium Rhodobacter sphaeroides
; FILE REFERENCE: CL2169 US NA
; CURRENT APPLICATION NUMBER: US/10/621.826
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1850
; TYPE: DNA
; ORGANISM: R. sphaeroides
US-10-621-826-2

Query Match		100.0%	Score 1850;	DB 7;	Length 1850;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1850;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	CCGAGGCGGCCCATCCAGAGATGCGCGCCGCTTTGCGCTGACGCCCGCGGATCTCTCGC	60		
Db	1	CCGAGGCGGCCCATCCAGAGATGCGCGCCGCTTTGCGCTGACGCCCGCGGATCTCTCGC	60		
Qy	61	TCGATCTGGGCGCCGCTGGGCGGACGATCGAAGCAAGAACTCTGATTAACCAAGCTTC	120		
Db	61	TCGATCTGGGCGCCGCTGGGCGGACGATCGAAGCAAGAACTCTGATTAACCAAGCTTC	120		
Qy	121	CACCTACCTTGGCGAGAGACAGAGAGTGAAGCAATGCTCGCCATGAGCCCCCGGAGC	180		
Db	121	CACCTACCTTGGCGAGAGACAGAGAGTGAAGCAATGCTCGCCATGAGCCCCCGGAGC	180		
Qy	181	CGGCGGTCGAGCTGATCGCCACATCGATCTGGACACAGGCCCATGCGCTGGCGAGCGCG	240		
Db	181	CGGCGGTCGAGCTGATCGCCACATCGATCTGGACACAGGCCCATGCGCTGGCGAGCGCG	240		
Qy	241	CGCGCGGATTTGCTTTGCCCTCCGCGCGCGACCGGTGCGTCCGTCGGAAGCGCGGC	300		
Db	241	CGCGCGGATTTGCTTTGCCCTCCGCGCGCGACCGGTGCGTCCGTCGGAAGCGCGGC	300		
Qy	301	TCGCGGCTGTCTATCCGCGAGCGGCCCATGTCTTAGGACTGACAAACCGGCTTCGGTCCC	360		

Db 301 |||||TCGCGCTGTATCCGCGAGCGCGCATGTCTACGGACTGACAACCGGCTCGGTCGCC 360
Qy 361 |||||TTGCGAACCGCTTGATCTCAGGTGAGAAATGTCGGAACGCTGCGAGCCAAATCTTGTCCATC 420
Db 361 |||||TTGCGAACCGCTTGATCTCAGGTGAGAAATGTCGGAACGCTGCGAGCCAAATCTTGTCCATC 420
Qy 421 |||||ATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGGAACGCGCGCGCCATCGTTCG 480
Db 421 |||||ATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGGAACGCGCGCGCCATCGTTCG 480
Qy 481 |||||CGGTCCTGGTTCGATCGCTCAGGAGCCTCCGGTGCACGAGAGGACCAATCGCTCGCC 540
Db 481 |||||CGGTCCTGGTTCGATCGCTCAGGAGCCTCCGGTGCACGAGAGGACCAATCGCTCGCC 540
Qy 541 |||||TGATCGACCTGCTCAATTCGAGCTCGCTCCGGCCGTTCCAGCGCGGACCGTGGGG 600
Db 541 |||||TGATCGACCTGCTCAATTCGAGCTCGCTCCGGCCGTTCCAGCGCGGACCGTGGGG 600
Qy 601 |||||CGTGGGTGACCTGACACCGCTTCGCGATATGCTGCTTCCTCCAGGGCCGGGAGACT 660
Db 601 |||||CGTGGGTGACCTGACACCGCTTCGCGATATGCTGCTTCCTCCAGGGCCGGGAGACT 660
Qy 661 |||||TCCTGGAACCGGAGACGGGCTTGACCGGCGAGAAAGGGCTCCGGCGCGGACGGCTGC 720
Db 661 |||||TCCTGGAACCGGAGACGGGCTTGACCGGCGAGAAAGGGCTCCGGCGCGGACGGCTGC 720
Qy 721 |||||AACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGGTCAACGGGACCTCCGCGCATGA 780
Db 721 |||||AACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGGTCAACGGGACCTCCGCGCATGA 780
Qy 781 |||||CCGGGATCGCGCTGGTGAATGCTCACGCCCTGCGCCCATCTCGGCAACTGCGGCGGTGCGT 840
Db 781 |||||CCGGGATCGCGCTGGTGAATGCTCACGCCCTGCGCCCATCTCGGCAACTGCGGCGGTGCGT 840
Qy 841 |||||TGAACGGCCCTGCTTGCGGAATGCTGAGAGCGCGGACCGAGGATGCGGCGCGGCACTGT 900
Db 841 |||||TGAACGGCCCTGCTTGCGGAATGCTGAGAGCGCGGACCGAGGATGCGGCGCGGCACTGT 900
Qy 901 |||||CCGACCTGCGGCGCATCCCGGACAGNAGACCGCGGAGGAGCTCGCGCGCGGTG 960
Db 901 |||||CCGACCTGCGGCGCATCCCGGACAGNAGACCGCGGAGGAGCTCGCGCGCGGTG 960
Qy 961 |||||ACGGCAGCGCGCGGTGCTCGGCACTGCTATTCGAGCGGAGGCTCGACGCGCGGATA 1020
Db 961 |||||ACGGCAGCGCGCGGTGCTCGGCACTGCTATTCGAGCGGAGGCTCGACGCGCGGATA 1020
Qy 1021 |||||TCGGGACGGAGCGGAGCGGGGACGATGCTTACAGCTTGCCTGCGCTCCGAGGTTTC 1080
Db 1021 |||||TCGGGACGGAGCGGAGCGGGGACGATGCTTACAGCTTGCCTGCGCTCCGAGGTTTC 1080
Qy 1081 |||||TCGGGCGGGCTTCGACACGCTCGCATGCGATGACCGGGTGTGACGATCGAGCTGAACG 1140
Db 1081 |||||TCGGGCGGGCTTCGACACGCTCGCATGCGATGACCGGGTGTGACGATCGAGCTGAACG 1140
Qy 1141 |||||CGGTGACCGGACAAATCCGGTGTTCGCGCCGATGCGAGCGTCCGCTGCAAGGGGCA 1200
Db 1141 |||||CGGTGACCGGACAAATCCGGTGTTCGCGCCGATGCGAGCGTCCGCTGCAAGGGGCA 1200
Qy 1201 |||||ATTTCATGGGCCAGCATGTCGCGCTGACGTCGATGCGCTTCGCCACGCGCGTCAACGTTTC 1260
Db 1201 |||||ATTTCATGGGCCAGCATGTCGCGCTGACGTCGATGCGCTTCGCCACGCGCGTCAACGTTTC 1260
Qy 1261 |||||TGGGGCGGCTTCGGAGCGGACGATTCAGCTCTGACGATGAAAGGCTGAAACGTTGGG 1320
Db 1261 |||||TGGGGCGGCTTCGGAGCGGACGATTCAGCTCTGACGATGAAAGGCTGAAACGTTGGG 1320
Qy 1321 |||||TGCCCCCTTCCTCCACCGGGGCGCGCGGGTTGAATTCGGCTTCATGGGCGCACAGG 1380
Db 1321 |||||TGCCCCCTTCCTCCACCGGGGCGCGCGGGTTGAATTCGGCTTCATGGGCGCACAGG 1380
Qy 1381 |||||TGAACGGGACCGCGCTCTCGGCGAGATGCGAGCCACGGGACCTGCTCGATCCATTCGA 1440

Db 1381 |||||TGACGCGGACCGCGCTCTCTGGCGGAGATCGAGCCACGGGACCTGCTCGATCCATTCGA 1440
Qy 1441 |||||TCTCCACGAACCGCGCAATCAGGATGTGCTTCGCTTGGGACCATCGCGCGCGCTCT 1500
Db 1441 |||||TCTCCACGAACCGCGCAATCAGGATGTGCTTCGCTTGGGACCATCGCGCGCGCTCT 1500
Qy 1501 |||||GCCGCGAGAAGATTCGACCGTTGGGCGGAGATCCTTCGATTCCTGCTCTGCTTCGAC 1560
Db 1501 |||||GCCGCGAGAAGATTCGACCGTTGGGCGGAGATCCTTCGATTCCTGCTCTGCTTCGAC 1560
Qy 1561 |||||AAGCTCGGAGCTGCGCTCGGCGAGGCTAGACGGGGTGTCTCCGCGGGGAAAGC 1620
Db 1561 |||||AAGCTCGGAGCTGCGCTCGGCGAGGCTAGACGGGGTGTCTCCGCGGGGAAAGC 1620
Qy 1621 |||||TGGTGCAGCGCTTCGCGGAGAGTTCCCGCGCTTGAGACGGACCGGCGCTCGGACAGG 1680
Db 1621 |||||TGGTGCAGCGCTTCGCGGAGAGTTCCCGCGCTTGAGACGGACCGGCGCTCGGACAGG 1680
Qy 1681 |||||AAATTGCGCGCTTGCTACGCACTCTTTCGAGCAATCTCCCGTTCGAGCGGCGCTCAGG 1740
Db 1681 |||||AAATTGCGCGCTTGCTACGCACTCTTTCGAGCAATCTCCCGTTCGAGCGGCGCTCAGG 1740
Qy 1741 |||||TGTGCGGGGATTCACCGGTGAGGCTTACAGGCTCGGCAATTTGAGCGAGGCTTCGATCG 1800
Db 1741 |||||TGTGCGGGGATTCACCGGTGAGGCTTACAGGCTCGGCAATTTGAGCGAGGCTTCGATCG 1800
Qy 1801 |||||ACGCCACGACGGCAGCGACCTCGATGGCGATCAGCTCGATCCGACGAGG 1850
Db 1801 |||||ACGCCACGACGGCAGCGACCTCGATGGCGATCAGCTCGATCCGACGAGG 1850

RESULT 2
US-11-069-633-7
; Sequence 7, Application US/11069633
; Publication No. US20050208643A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt-Dannert, Claudia
; APPLICANT: Watte, Kevin
; TITLE OF INVENTION: Flavonoids
; FILE REFERENCE: 09531-160001
; CURRENT APPLICATION NUMBER: US/11/069,633
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: US 60/549,077
; PRIOR FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1581
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-11-069-633-7

Query Match 85.4%; Score 1580; DB 10; Length 1581;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 TGAAGCCAATGCTCGCCATGAGCCCCCGAAGCCGSCCGTTCGAGCTGGATCGCCACATCG 207
Db 2 TGAAGCCAATGCTCGCCATGAGCCCCCGAAGCCGSCCGTTCGAGCTGGATCGCCACATCG 61
Qy 208 ATCTGGACCAAGCCCATGTCGCTGGCGAGCGGCGCGGATTTGCTTGGCCCTCCGG 267
Db 62 ATCTGGACCAAGCCCATGTCGCTGGCGAGCGGCGCGGATTTGCTTGGCCCTCCGG 121
Qy 268 CGCGCGACCGGTGCGCTGCGAAGCGCGGCTCGGCGCTGTTCATCCGCGGCGCGCC 327
Db 122 CGCGCGACCGGTGCGCTGCGAAGCGCGGCTCGGCGCTGTTCATCCGCGGCGCGCC 181
Qy 328 ATGTCTACGAGCTGACCAACCGGCTTCGGTCCCTTCGGAACCGCTGTATCTCAGGTGAGA 387
Db 182 ATGTCTACGAGCTGACCAACCGGCTTCGGTCCCTTCGGAACCGCTGTATCTCAGGTGAGA 241
Qy 388 ATGTCCGAACCGCTGACGGCCAATCTTGTCCATCATCTGGCCAGCGGCGTGGGACCGGTGC 447

```
Db 242 ATGTCGGAACGCTGACGGCAATCTTGTCCATCATCTGGCCAGCGCGTGGGACCGGTGC 301
Qy 448 TTGACTGACAGACGGCGCGCATGGTTCTGGCCGCTGTGTGTCATGCTCAGGAG 507
Db 302 TTGACTGACAGACGGCGCGCATGGTTCTGGCCGCTGTGTGTCATGCTCAGGAG 361
Qy 508 CCTCCGGTGCAGACGAGGGACCATCGCTCCGCTGATCGACTGCTCAATTCGAGCTCG 567
Db 362 CCTCCGGTGCAGACGAGGGACCATCGCTTCGCTGATCGACTGCTCAATTCGAGCTCG 421
Qy 568 CTCGGCCGTTCCACAGCCGCGACGGTGGGCGCTCGGGTGACCTGACACCGCTTGGCG 627
Db 422 CTCGGCCGTTCCACAGCCGCGACGGTGGGCGCTCGGGTGACCTGACACCGCTTGGCG 481
Qy 628 ATATGGTGCTCTGCTCTCAGGGCCGGGAGACTTCTCTGGACCGGGACCGGGCTTGG 687
Db 482 ATATGGTGCTCTGCTCTCAGGGCCGGGAGACTTCTCTGGACCGGGACCGGGCTTGG 541
Qy 688 ACGGCGAGAGGGCTCCGGCGCGGACGGCTGCAACCGCTCGATCTCTCCATCCGATG 747
Db 542 ACGGCGAGAGGGCTCCGGCGCGGACGGCTGCAACCGCTCGATCTCTCCATCCGATG 601
Qy 748 CACTGGCGCTGCTCAACGGGACCTCCGCGATGACCGGGATCGCGTGGTGAATGCTCAG 807
Db 602 CACTGGCGCTGCTCAACGGGACCTCCGCGATGACCGGGATCGCGTGGTGAATGCTCAG 661
Qy 808 CCTGCCGCGCATCTCGGCAACTGGGGCGGTGGCGTTGACCGGCCCTGCTTGGCGAATGCTCGA 867
Db 662 CCTGCCGCGCATCTCGGCAACTGGGGCGGTGGCGTTGACCGGCCCTGCTTGGCGAATGCTCGA 721
Qy 868 GAGGCGGACCGAGCATGGGCGCGGACACTGTGCGACCTCGGCGGATCCCGGACGAGA 927
Db 722 GAGGCGGACCGAGCATGGGCGCGGACACTGTGCGACCTCGGCGGATCCCGGACGAGA 781
Qy 928 AGGACGCGCAGCAGGCTGCGCGCCCGCTGGAGCGGACGCGCGGTGTCTCGGCACG 987
Db 782 AGGACGCGCAGCAGGCTGCGCGCCCGCTGGAGCGGACGCGCGGTGTCTCGGCACG 841
Qy 988 TCATTGGCGAGCGGAGCTCGACGCGCGCGATATCGGGACCGGACCGGAGCGGGCGAGG 1047
Db 842 TCATTGGCGAGCGGAGCTCGACGCGCGCGATATCGGGACCGGAGCGGGCGAGG 901
Qy 1048 ATGCTACAGCCTGGCTGCGCTCGGAGGTCTCGGGCGGGCTTGGACAGCTCGAT 1107
Db 902 ATGCTACAGCCTGGCTGCGCTCGGAGGTCTCGGGCGGGCTTGGACAGCTCGAT 961
Qy 1108 GGCATGACCGGGTGTGACGATCGAGCTGAACGGGTGACCGGATCCCGGTGTTCCGC 1167
Db 962 GGCATGACCGGGTGTGACGATCGAGCTGAACGGGTGACCGGATCCCGGTGTTCCGC 1021
Qy 1168 CCGATGGCAGCGTCCCGCCCTGACAGGGGGCAATTTTTCATGGGCGAGCATGTGGCGCTGA 1227
Db 1022 CCGATGGCAGCGTCCCGCCCTGACAGGGGGCAATTTTTCATGGGCGAGCATGTGGCGCTGA 1081
Qy 1228 CGTCCGATGCGCTCGCACGCGCGTCAACGTTCTGGGGGGCTTGGGAGCGCCAGATTG 1287
Db 1082 CGTCCGATGCGCTCGCACGCGCGTCAACGTTCTGGGGGGCTTGGGAGCGCCAGATTG 1141
Qy 1288 CACCTCTGACAGATGAAGGCTGAACGTTGGGCTGCCCGCTTCTTCCACCGGGGGCCCG 1347
Db 1142 CACCTCTGACAGATGAAGGCTGAACGTTGGGCTGCCCGCTTCTTCCACCGGGGGCCCG 1201
Qy 1348 CCGGGTTGAATTCGGGCTTTCATGGGCGCACAAGGTGAAGGGACCGCGCTCTCTGGCCGAGA 1407
Db 1202 CCGGGTTGAATTCGGGCTTTCATGGGCGCACAAGGTGAAGGGACCGCGCTCTCTGGCCGAGA 1261
Qy 1408 TGCAGGCCACGGGACCTGCTCGATTCGATTCGATTCGATTCGAAACCGCGCCCAATCAGATTG 1467
Db 1262 TGCAGGCCACGGGACCTGCTCGATTCGATTCGATTCGATTCGAAACCGCGCCCAATCAGATTG 1321
Qy 1468 TGCTCTGCTTGGGACCATCGCGCGGCTCTCGCGGAGAGATCTCTGGACCGGACCGGCTTGAGCGG 1527
```

```
Db 1322 TGCTCTGCTTGGGACCATCGCGCGGCGCTCTGCGCGAGAGATCGACCGTTGGCGG 1381
Qy 1528 AGATCTTGGCATCTCGCTCTCTGTCTTGTGCAAGCTGCGGAGCTGCGTGCAGCG 1587
Db 1382 AGATCTTGGCATCTCGCTCTCTGTCTTGTGCAAGCTGCGGAGCTGCGTGCAGCG 1441
Qy 1588 GCCTAGACGGGCTGCTCCCGGGGAGAGAGCTGTGTCAGGCCCTGCGGAGCAGTTCC 1647
Db 1442 GCCTAGACGGGCTGCTCCCGGGGAGAGAGCTGTGTCAGGCCCTGCGGAGCAGTTCC 1501
Qy 1648 CGCGCTTGGACGAGCGCGGCCCTTGGGACAGGAAATTCGCGCGCTTGTACGCACTCT 1707
Db 1502 CGCGCTTGGACGAGCGCGGCCCTTGGGACAGGAAATTCGCGCGCTTGTACGCACTCT 1561
Qy 1708 TGCAGCAATCTCCGCTCGA 1727
Db 1562 TGCAGCAATCTCCGCTCGA 1581

RESULT 3
US-10-369-493-31503
; Sequence 31503, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31503
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31503

Query Match 78.2%; Score 1446; DB 6; Length 1446;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 273 GACCGGTGCGGTGCGTCCGAGCGCGCTCGGCGCTGTCATCCGAGCGCGCATGTC 332
Db 1 GACCGGTGCGGTGCGTCCGAGCGCGCTCGGCGCTGTCATCCGAGCGCGCATGTC 60
Qy 333 TACGGACTGACAAACCGGCTTCGGTCCCTTTCGCAACCGCCCTGATCTCAGGTGAGAATGTC 392
Db 61 TACGGACTGACAAACCGGCTTCGGTCCCTTTCGCAACCGCCCTGATCTCAGGTGAGAATGTC 120
Qy 393 CGAACGCTGAGGCAATCTTGTCCATCATCTGGCCAGCGCGTGGGACCGGTGTTGAC 452
Db 121 CGAACGCTGAGGCAATCTTGTCCATCATCTGGCCAGCGCGTGGGACCGGTGTTGAC 180
Qy 453 TGGACGAGCGCGCGCATGGTTCTTGGCGGTCTGCTGATCGCTCAGGAGCGCTCC 512
Db 181 TGGACGAGCGCGCGCATGGTTCTTGGCGGTCTGCTGATCGCTCAGGAGCGCTCC 240
Qy 513 GGTGCCAGCGAGGGGACCATCGCTCGCTCGATCGACCTGCTCAATTCGAGCTCGCTCCG 572
Db 241 GGTGCCAGCGAGGGGACCATCGCTCGCTCGATCGACCTGCTCAATTCGAGCTCGCTCCG 300
Qy 573 GCGCTTCCGAGCGCGGACCGGTGGGCGCGTGGGTGACCTGACACCGCTTGGCGATATG 632
Db 301 GCGCTTCCGAGCGCGGACCGGTGGGCGCGTGGGTGACCTGACACCGCTTGGCGATATG 360
Qy 633 GTGCTCTGCTTCCAGGGCGGGGAGACTTCTTGGACCGGAGCGGACCGGCTTGAGCGG 692
```

```
Db 361 GTGCTCTCTCCAGGGCCGGGAGACTTCTCTGACCGGGACGGGACGGCGCTTGACGGC 420
Qy GCAGAAGGGCTCCGGCGGGAGCGCTCAACCGCTCGATCTCTCCCATCGGATGCACTG 752
Db 421 GCAGAAGGGCTCCGGCGGGAGCGCTCAACCGCTCGATCTCTCCCATCGGATGCACTG 480
Qy 753 GCGCTGCTCAACGGGACCTCCGCATGACCGGGATCGCGTGGTGAATGCTCAGCCCTGC 812
Db 481 GCGCTGCTCAACGGGACCTCCGCATGACCGGGATCGCGTGGTGAATGCTCAGCCCTGC 540
Qy 813 GCGCATCTCGGCAACTGGCGGTGGCTGAGAGCCCTGCTTGGCGAATGCTGAGAGGC 872
Db 541 GCGCATCTCGGCAACTGGCGGTGGCTGAGAGCCCTGCTTGGCGAATGCTGAGAGGC 600
Qy 873 CGGACCGAGGATGGGCGGGGCACTCTCGACCTGGGGCGGATCCCGGACAGAGGAC 932
Db 601 CGGACCGAGGATGGGCGGGGCACTCTCGACCTGGGGCGGATCCCGGACAGAGGAC 660
Qy 933 GCGCAGCGAGGCTCGCGCCCGCTGGAGCGGCGGGGTGGTCCGGCAGCTCATTT 992
Db 661 GCGCAGCGAGGCTCGCGCCCGCTGGAGCGGCGGGGTGGTCCGGCAGCTCATTT 720
Qy 993 GCGGAGCGGAGGCTCGAGCGCGGGCATATCGGAGCGGAGCGGGGCGAGGATGCC 1052
Db 721 GCGGAGCGGAGGCTCGAGCGCGGGCATATCGGAGCGGAGCGGGGCGAGGATGCC 780
Qy 1053 TACAGCTCGGCTCGCGCTCCGCAAGTTCTCGGGCGGGCTTCGACACGCTCGCATGGCAT 1112
Db 781 TACAGCTCGGCTCGCGCTCCGCAAGTTCTCGGGCGGGCTTCGACACGCTCGCATGGCAT 840
Qy 1113 GACCGGTGCTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGGTGTTCGCGCCGAT 1172
Db 841 GACCGGTGCTGACGATCGAGCTGAACCGCGGTGACCGCAATCCGGTGTTCGCGCCGAT 900
Qy 1173 GCGAGCGTCCCGCTTCGCAAGGGGCAATTTATGGGCGAGATGCGCTGAGCTGC 1232
Db 901 GCGAGCGTCCCGCTTCGCAAGGGGCAATTTATGGGCGAGATGCGCTGAGCTGC 960
Qy 1233 GATGCGCTCGCAGGGCGCTCACGTTCTCGCGGGCTTCGGGAGCGCCAGATGGACGT 1292
Db 961 GATGCGCTCGCAGGGCGCTCACGTTCTCGCGGGCTTCGGGAGCGCCAGATGGACGT 1020
Qy 1293 CTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTTCCACCGGGGCGCCCGCGG 1352
Db 1021 CTGACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCTTCCACCGGGGCGCCCGG 1080
Qy 1353 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGACCGCGTCTCTGGCGGAGATGCGA 1412
Db 1081 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGACCGCGTCTCTGGCGGAGATGCGA 1140
Qy 1413 GCCACGGGACCTGCTCGATCCATTGATCTTCCAGAACCGCGCCATCAGGATGGTTC 1472
Db 1141 GCCACGGGACCTGCTCGATCCATTGATCTTCCAGAACCGCGCCATCAGGATGGTTC 1200
Qy 1473 TCGCTTGGGACCATCGCGCGGCTCTGCGGAGAGATGCAACCGTTTGGGCGGAGATC 1532
Db 1201 TCGCTTGGGACCATCGCGCGGCTCTGCGGAGAGATGCAACCGTTTGGGCGGAGATC 1260
Qy 1533 CTTGCGATTCCTCGTCTCTGTTGCAAGCTTGGAGCTGCGCTGCGGAGCGGCGCTA 1592
Db 1261 CTTGCGATTCCTCGTCTCTGTTGCAAGCTTGGAGCTGCGCTGCGGAGCGGCGCTA 1320
Qy 1593 GACGGGTGCTCCCGGGGAGAGAGCTGTCAGGCGCCCTGCGGAGAGTTCGCGCG 1652
Db 1321 GACGGGTGCTCCCGGGGAGAGAGCTGTCAGGCGCCCTGCGGAGAGTTCGCGCG 1380
Qy 1653 CTTGAGACGAGCGGCGCCCTGGGACAGAAATTCGCGCTTGTGACGACCTCTTTCGAG 1712
Db 1381 CTTGAGACGAGCGGCGCCCTGGGACAGAAATTCGCGCTTGTGACGACCTCTTTCGAG 1712
Qy 1713 CAATCT 1718
Db 1441 CAATCT 1446
```

RESULT 4

```
US-10-464-609-3
; Sequence 3, Application US/10464609
; Publication No. US20040029230A1
; GENERAL INFORMATION:
; APPLICANT: KYNDT, John, Jozef Armand
; APPLICANT: VAN BEEUMEN, Jozef
; TITLE OF INVENTION: No. US20040029230A1el Methods For Synthesis of
; FILE OF INVENTION: Holo-Photoactive Yellow Protein
; FILE REFERENCE: 50304/008001
; CURRENT APPLICATION NUMBER: US/10/464,609
; CURRENT FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/389,593
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Rhodobacter capsulatus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1623)
; OTHER INFORMATION: n = A,T,C or G
US-10-464-609-3
```

```
Query Match 30.3%; Score 561; DB 7; Length 1626;
Best Local Similarity 60.9%; Pred. No. 2.6e-137;
Matches 947; Conservative 0; Mismatches 603; Indels 6; Gaps 2;

Qy 175 CGAAGCGGCGCTCGAGCTGGATCGCCACATCATCTGGACGAGGCCATGCGTGGCGA 234
Db 53 CCAAGACTGCTCGCGCTGGACGGGGCGCTGACATTTGTCCAATGAGGCCATCGCA 112
Qy 235 GCGGCGCGCGCGGATTTGCTTTCCTTCGCGCGCGCGACCGGTGCGTCCGAAG 294
Db 113 CACATCGCAGCCGATTTCCGTGACCCCGCGCTGCGGAGCGCTGCGCGGGCCATG 172
Qy 295 CGCGCTTCGCGCTGTTCATCGCGAGCGCGCCATCTTACGAGCTGACAAACCGGCTTCG 354
Db 173 CCGCGCTTGAGCAGCGCATCGCGCAGCAGCGCCACATTTACGGCATCACACCGGCTTCG 232
Qy 355 GTCCCTTCGGAACCGCTGATCTCAGGTGAGAACTCCGAAACGCTGCAGGCAATCTTG 414
Db 233 GCCGCTGGGCAACCGCTGATCGGGGCGCATCAGGGGCGGAGCTGCAGAGAACCTGA 292
Qy 415 TCATCATCTGGCAGCGCGTGGGACCGGTGCTTGAATGAGACGAGCGCGCGCCATGG 474
Db 293 TCTATCATCTGGCCACCGGGCTCGGCGCGAAACTGAGCTGGGCGGAGCGCGGGCTTGA 352
Qy 475 TTCTGGCGGTCTGGTGTGATCGCTCAGGAGCGCTCCCGTTCGCGAGGGGAGCCATCG 534
Db 353 TGCTGGCGGTCTCAACTCGATCTCTGCAAGGCGCTCGGGGCGCTCGCCGAGACGATCG 412
Qy 535 CTGCGCTGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCACAGCGCGGCAACCG 594
Db 413 ACCGATTCGTTGGGTGCTCAATTCGCGGGTTCGCGCGAGGTTCCGGCGCAGGAAACCG 472
Qy 595 TGGGCGCTCGGTGATGATGACACCGCTTCGCGCATATGTTGCTCTGCTCCAGGGCGCGG 654
Db 473 TGGGCGCTCGGTGATGATGACACCGCTTCGCGCATATGTTGCTCTGCGGCTGCAGGAGCGG 532
Qy 655 GAGACTTCTTGAACCGGAGCGGACCGGGCTTGAACCGCGCAGAGGGCTCCGGCGCGGAC 714
Db 533 GCGCGATGATCGACCCCTCGGGCGCGGTGCGAGAGGCGCGGGCGGTGATGGATCGGCTCT 592
Qy 715 GGCTGCAACCGCTCGATCTCTCCCATCGGATGACCTTGGCGCTGTTCAACGCGGACCTCG 774
Db 593 GCGGCGGTTCGCTGACGCTGGCGCGCGGTGACGCGGCTGGCGCTGGGCTGATGGACCTCGG 652
Qy 775 CCATGACCGGGATCGCGCTGGTGAATGCTCAGCGCTTCGCGCCATCTCGGCAACTCGGCGG 834
```

Db 653 CGATGACCGGATTTGGGCGCTGACCGGGTCTGAGGCGCGGCGGCGATCGACCGCGC 712
Qy 835 TGGCGTTGACCGCCCTGTTCCGGAATGTTGAGAGCGCGGACCGAGGCTATGGCGCGG 894
Db 713 TTCCGACAGCGCGGTCTGATGAGGTCTTGTCCGGTCTATGCGGAAGCCTGGCATCCG 772
Qy 895 CACTGTCCGACCTCGCGCGCATCCCGACAGAGAGCGCGCAGCGAGGCTGCGCGCC 954
Db 773 CTTTCCGAGAGCTCGCGCGCATCCCGGCGAGTCTGCGGCGACCGAGCGGTGCGGAGG 832
Qy 955 GCGTGGACGCGCAGCGCGGGTGTCTCGGCACGCTCATTTGCCGAGCGAGGCTCGACCGCG 1014
Db 833 CGCTGGATGGCGCGGGCGGTCTGTGCGACCTTGACCGCGCGCGGGCTGACCGCG 892
Qy 1015 GCGATATCGGACCGGACCGGAGCGCGGGGCAAGATGCTTACAGCCTCTGCGCTCGC 1074
Db 893 CGGATCTCGCGCGCGGAGATCATCCGCGCAGAGTGTCTACGATCTCGCGGTGTGCGCG 952
Qy 1075 AGGTTCTCGGGCGGGCTTCGACAGCTCTCGCATGCGATGACCGGGTGTGACGATCGAGC 1134
Db 953 AACTGGTCTCGCGCGGTCTGCGACACGCTGGACCTGGACCTGCTGCTCGTCACTGCGAGC 1012
Qy 1135 TGAACGCGGTGACCGACAACTCCGGTGTTCGCGCCGATGGCAGCGTGC CGCGCCCTGCA 1194
Db 1013 TCAATTCGTACCGACAACTCCGATCTTCCGAGGGCTGCGCGGTGCGCGCCCTGCA 1072
Qy 1195 GGGCAATTTATGGCGCAGCATGTGCGTGTGACGTCCGATGCGCTCGCCACCGCGCTCA 1254
Db 1073 GCGCAATTTATGGCGGTGCTGCGACAGCTGCGTCCGCTGCGTCAACCGCGCGCTG 1132
Qy 1255 CCGTTCGCGGGCTTCGCGAGCGCGCAGATGCGATGCGCTGCGACAGTGAAGGCTGAAC 1314
Db 1133 TGAAGTGGCGGGCTGCTGCGAGCGTCAATCGCGCGCTGACCGACGAAAGCTGAACA 1192
Qy 1315 GTGGCTGCGCCCTTCTCCACCGCGCGCGCGCGGTGAAATTCGCGCTTCATGGCG 1374
Db 1193 AGGCTTGC CGCGCTTCTGCA CGGGGGGCGAGCGGGGCTGCAATCGCGCTTCATGGGG 1252
Qy 1375 CACAGGTGACGGGACCGCGCTCTGCGCGAGATGCGAGCCA ---CGGACCTGCTCGA 1431
Db 1253 CGCAGGTCA CGCGGACGCGCTTCTGCGGAAATGCGGGGAGATGCGACCGCGTTCG 1312
Qy 1432 TCAATTCGATCTCACGAAACCGCGCAATCGAGATGCTGCTGCGTGGGACCAATCGCG 1491
Db 1313 TGCAGTCTGCTGCAACCAATGCGCGCAATCAGGATGTGCTGATGCGAACGATTCGCG 1372
Qy 1492 CGCGCTCTGCGCGAGAGATCGACCGTTGGCGGAGATCCTTGGCATCTCGCTCTCT 1551
Db 1373 CGCGAGGGCGCGGCGGAGCTGCTGCCCTGTGCGAGATCCAGGCGATCTGCGCGTTG 1432
Qy 1552 GTCTTGACAAAGCTGCGGAGCTGCGCTGCGGCGAG ---GGCTAGACGGGGTGTCTCCG 1608
Db 1433 CCCTTGCCAGGCGATGATCTGTTGACGACCCCGAGGGGCGCGGATGCTGCTTA 1492
Qy 1609 CGGGGAAGAGTGTGTGAGCGCCCTGCGGAGAGATGTTCCGCGCTTGAAGAGGACCGCG 1668
Db 1493 CGCGCGGGATCTGCGGAGCGGATCGCGGCGGTCTGCGCGGGCTTCGCGCGCGACAGAC 1552
Qy 1669 CCCTGGAGAGGAATTCGCGCTGCTAGCACCTCTTGCAGCAATCTCCCGTC 1724
Db 1553 CGCTTCCGGGATATCGAAGCGGTGCGACAGGGTCTGCGCTCCCTCCNNNGCC 1608

RESULT 5

US-10-369-493-32499
; Sequence 32499, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 32499
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Ralstonia metallidurans
US-10-369-493-32499

Query Match 13.2%; Score 244.2; DB 6; Length 1587;
Best Local Similarity 50.2%; Pred. No. 3.9e-54;
Matches 780; Conservative 0; Mismatches 743; Indels 30; Gaps 6;

Qy 178 AGCGGCGCTCGAGCTGCGGCACATCGATCTGGACACAGGCCCATGCGGTGGGAGCG 237
Db 2 ATCGGCGCGACATCGACCGCCACACCTGACCCCGGACCGTCCGCGCCCATCGACGCG 61
Qy 238 GCGGCGCGGATTGTCTTCCCTCCGCGCGCGACCGGTGCGGTCCGTCGTCGGAAGCG 297
Db 62 GCAGCGCGCGGCATCGTCCGAGCCCGTCTCGGCAAGGTTCGCGATGCCCGGCC 121
Qy 298 GGCTCGGCGCTGTCTATCCGCGAGCGCGCATGTCTACGAGTGAACAACCGGCTTCGCTC 357
Db 122 GCTTCGAGCAGGTGGCTGCGGCCAAATGTGCGCATCTACGGCGTCTCCACGGGCTTGGCG 181
Qy 358 CCCTTGGGAACCGGCTGATCTCAGGTGAGAAATGTCCGAACTGTCGACGCGCCATCTTGTCC 417
Db 182 AACTGTGTACAACTGGGTGCGATCGAATCGAATGCGCGTGGCGTGCAGAGAGAACCTGTTC 241
Qy 418 ATCATCTGGCGCAGCGGCGTGGGACCGGTGTCTGACAGACGCGCGCGCGCCATGTTTC 477
Db 242 GCAGCCATTGCGGGGTGTTGGTCCGCTGTTTCGCGGAGAGGTTCGCGCGATGATGG 301
Qy 478 TGGCGGCTGTGTGTGATCGCTCAGGAGCCTCCGCTGCGAGCGAGGGGACCATCGCTC 537
Db 302 TC GCGGCTGCAATCGCTTGGCACCGGATCTCGCGGTGCGCGCGCGTTCGGAAC 361
Qy 538 GCCTGATGACGCTGCTCAATTCGAGTCTCGTCCGCGCGTTCGAGCGCGCGGACCGTGG 597
Db 362 AACTGTGAAATGATCTGGAAGCGGCAATCACGCGCGCGTGCAGAGTGGGTTTCGCTCG 421
Qy 598 GCGGTCGGGTGACCTGACACCGCTTCCGATATGTTGCTCTGCTCCAGGGCGGGAG 657
Db 422 GGGCAGCGGTGATCTCGCGCTCTGTGCGCATCTGCTGCGCATCACTGCTGATCGCGAAGCA 481
Qy 658 ACTTCTCTGACCGGGACCGGACCGGCTTGAACCGCGCAGAGGGCTCCGCGCGGACCGG 717
Db 482 AGGTGCTGACCGAGGATGGCGGTACGCCACCCACGCGCGGAAGTCTCTGCGGAGGTGGCA 541
Qy 718 TGCAACCGCTGCTGATCTCTCCATCGCGATGCACTGGCGCTGGTCAACGAGGACCTCCGCA 777
Db 542 TCACGCGCTCGCGCTGCGCTGACAAAGGAGGGCTGCGCTGATCAACGAGGACATCGGCCA 601
Qy 778 TGACCGGGATCGCGCTGGTGAATGCTCACGCTGCGCGCATCTCGGCACTGGGCGGTGG 837
Db 602 TGACCGGGGTGTGTGCTGCTGTTGAGAGCGTGTGCGCGCGCAGGTCCAGAGGCCGAGA 661
Qy 838 CGTTGACGCGCTTGTGCGGAATGTCTGAGAGCGCGGACCGAGGCAAT ---GGCGCGCG 894
Db 662 TCATCGGGCGCTGGCGCTCGAAGGATATCCCGCTCGCGCATGCTTCTATGGCCCATG 721
Qy 895 CACTGTCCGACCTGCGCGCGCATCCCGGACAGAGGACGCGCGAGCGAGGCTGCGCGCC 954
Db 722 GGCAAGACATCGCCAAACCGCATCCGGAACAGATCCGCTCGCGCGGCAACATGCGCGCGC 781
Qy 955 GCCTGGACGCGCGCGGTGTTCCGCGACG -----TCATTCGCGGAGGGGCTCG 1008

```
Db 782 TGCTGGCCGATTCGGACACGGCTCTCCGACATGGCGAACTGTCCGCGAGATGAAGACAC 841
Qy 1009 ACGCCGCCGATATCGGACACGAGCCGAGGCGGG-----CAGGATGCCCTACAGCCTGC 1062
Db 842 GCGGGGCGAGGCAAGAACCGGCACTGGCGTTCATCCAGAGGCCCTACACGCTGC 901
Qy 1063 GCTGCGCTCCGAGGTTCTCGGGCGGGCTTCGACACGCTCGCATGCGATGACCGGGTGC 1122
Db 902 GCTGCATTCGCGAGGTGCTTGGCGGGTGGCGATACGCTCGACCATTCGCGCACCGTGG 961
Qy 1123 TGACGATCGAGCTGAACCGGTGACCGACAAATCGGTGTTTCCGCCGATGGCAGCGTGC 1182
Db 962 TCGAGCGCGAATGAATTCATCGAATGACAATCGCTGTTCTTCGAAGACGGCGAGCTGT 1021
Qy 1183 CCGCCCTGACCGGGGCAATTTTCATGGCCAGCATGTGGCGCTGAGCTCCGATGCGCTCG 1242
Db 1022 TC-----CAGGGGGCACTTTCACGGCCAGCGAGTGGATTTCGCAATGACTTCTTGG 1075
Qy 1243 CCACGGCGGTCA CGGTTCTGGCGGGCTTTCGGAGCGCCAGATTGACCGTCTGACAGATG 1302
Db 1076 CCATCGCGGCA CGCAACTGGGGGTGGTGTGGAGCGCGCTGAACCGGCTCTGTGAGCC 1135
Qy 1303 AAAGGTGAACCGTGGGTGCCCCCTTCTTCCACCGGGCCCCGGGGTTGAATTCGG 1362
Db 1136 CGCATCTGAACAACAATCTGCGGGCGTTCTGCGGGCGGAAACGAGGGGTTTCTGCTGG 1195
Qy 1363 GCTTTCATGGCGCACAGGTGACGGCGACCGCTCTCTGGCGGAGATGCGAGC---CACGG 1419
Db 1196 GGTTCGGGGGACAGTATCCGGCCACGGCGTGTGATTCGGAGAACCGCACGATGTGCA 1255
Qy 1420 GACCTGCTCGATCCATTCGATTCACACGAACCGCCCAATCAGGATGTGTCTCGCTTG 1479
Db 1256 GCCCGCGAGCATCCAGAGTGTCCGTCGAACGGCGACACACAGGATGTGTTCAGCATGG 1315
Qy 1480 GGACCATCGCGCGCGCTTTCGGCGGAGAAATGACCGTTGGGGCGGAGATCCTTCGGA 1539
Db 1316 GGCTGATCGCTGCCCGCAATTCGCCCGCGCATTCGACCAATAACAGTATCATCTCGCGC 1375
Qy 1540 TCCTCGCTCTCTCTTTCGACAACTGCGGAGTGTGCTGCGGCGCGGCTAGACGGGG 1599
Db 1376 TGGAGTTGCTGGCGTCATGTACGCGCGCGAAT-----CGCGGGCGCGGTTCGAGCAAC 1429
Qy 1600 TGTCTCCCGCGGGAAGAAGCTGGTCAGGCCCTGCGCGAGCAGTTCCCGCCGCTTGAGA 1659
Db 1430 TGGCGCGCGGAGCCCGCGCTGTTCGCTGCTGGGGAGCGCGTGGCGTTCCTCTCGA 1489
Qy 1660 CGACCGGCCCCCTGGGACAGGAAATTCGCGCGCTTGTCTAGGACCTCTTCGAG 1712
Db 1490 TCGATCGCTATATGACCGGACGACATCGAGGCTATGGCGGCTGCTCCGCTCAG 1542
```

```
RESULT 6
US-10-159-257A-1/c
; Sequence 1, Application US/10159257A
; Publication No. US20040161828A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, BEN
; APPLICANT: LIU, WEN
; APPLICANT: CHRISTENSON, STEVEN D.
; APPLICANT: STANDAGE, SCOTT
; TITLE OF INVENTION: GENE CLUSTER FOR PRODUCTION OF THE ENEDIYNE ANTITUMOR
; FILE REFERENCE: 407T-896020US
; CURRENT APPLICATION NUMBER: US/10/159,257A
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 207
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 42000
```

```
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5982)..(7475)
; OTHER INFORMATION: orf(-2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13012)..(14076)
; OTHER INFORMATION: orf10
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15901)..(15960)
; OTHER INFORMATION: splice variant a
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25815)..(27167)
; OTHER INFORMATION: orf20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27214)..(28590)
; OTHER INFORMATION: orf19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29632)..(31194)
; OTHER INFORMATION: orf2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31280)..(32587)
; OTHER INFORMATION: orf3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (32809)..(34389)
; OTHER INFORMATION: orf4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (36165)..(37487)
; OTHER INFORMATION: orf5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (37559)..(38935)
; OTHER INFORMATION: orf23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (38983)..(39261)
; OTHER INFORMATION: orf26
; US-10-159-257A-1

Query Match 13.1%; Score 242.4; DB 7; Length 42000;
Best Local Similarity 50.3%; Pred. No. 1.1e-53;
Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;

Qy 227 CGTGGCGAGCGGCGCGCGGATTCCTTGGCCCTCCGGCGCGGACCGGTGCGGTGC 286
Db 40906 CGTCGCGGAGGAACGCGCGACGCTGACGTACCGGCCGAATCCATCGGAAGCCCGAGAA 40847
Qy 287 GTCCGAAGCGCGCTCGGGCTGTCATCCCGAGGGCGGCCATGCTACGGACTGACAAC 346
Db 40846 GAGCCGGGAGATCTTCGAAGGGATCCCGAACAGAACATCCCCATCTACGGGTGACCAC 40787
Qy 347 CGGCTTCGTCCTCCCTTCGGAACCGCTGATCTCAGGTGAGAATGTCCGAACGCTGCAGGC 406
Db 40786 CGGGTACGCGAGATGATCTACATGCAAGTTCGACAAAGTGAAGGAGTGAAGTGCAGAC 40727
Qy 407 CAATCTTGTCCATCATCTGGCCAGCGGCGTGGGACCGGTGCTTGTGACTGACGACGCGGCG 466
Db 40726 CAATCTCGTCCGTAGCACACAGCGCGGGAGTCCGTCCGTGTTTCGCCGAGGACGAGCGCG 40667
Qy 467 CGCCATGTTCTTGGCGGCTGCTGTCGATCGCTCAGGAGAGCTCCGCTGCCAGCGAGGG 526
Db 40666 GCGGATCGTGGCGGCCCGGCTGAACACCTCGCAAGGGCCACTCCGCGGTGCGCCCAT 40607
Qy 527 GACCATCGCTCGCTGATCGACCTGCTCAATTCCGAGCTCGCTCGGGCGCTTCCAGCGCG 586
```

Db 40606 CATCTCGAACCCCTCGCGAGTACCTGAACAGGGCATCACCCGGCCATACCCGAGAT 40547
Qy 587 CGGACCGGTGGCGGTGGGTGACCTGACACACCGGTTGGCGATATGGTGTCTGCTCCCA 646
Db 40546 CGGGTCACTCGGGGCGAGCGGCGACCTGGCTCCCTCTCCCACTCGCGAGCACCTCAT 40487
Qy 647 GGGCCGGGAGACTTCTTGGACCGGAGCGGACCGGCTTGACCGCGACAGAGGCTCCG 706
Db 40486 CGGAGAGGCTACGTCTCTG---CGGACGGACGGCGGTGGAGACCGCCAGGTGCTGGC 40430
Qy 707 GCGCGAGCGGTGCAACCGCTCGATCTCTCCCATCGGATGCACTGGCGCTGGTGTCAACGG 766
Db 40429 CGACGGGSCATCGACCGCTCGAAGTGGCTTCAAGAGGGCTCGCACTGATCAACGG 40370
Qy 767 GACCTCGGCATGACCGGAGTGGCGGTGGTGAATGCTTCAACCGCTTGGCGCCATCTCGGCAA 826
Db 40369 CACGTCCGGATGACCGGTCTGGGCTCCCTGGTCTGGACCGGCGCTGGAGCAGGCCA 40310
Qy 827 CTGGCGGTGGCTTGGACCGGCTCTTGGGATGTCTGAGAGCGGACCGAGGCGATG 886
Db 40309 GCAAGCGAGATCGTGAAGCTCTGCTCATCGAGCGGTACGCGGATCGACACCGCCCTT 40250
Qy 887 GCGCGCGGCACTGTCGACCT---GCGGCGCATCCCGGACAGAGAGCGCGCAGCGAG 943
Db 40249 CCTCGGAGGGGACGACATAGCCCGCCGACAGGCGCCAGATCGACACCGCGCCAA 40190
Qy 944 GTCGCGCGCGGTGGACGCGCGGTGGTCTCGGACGTCATTTGCGGACGCGAG 1003
Db 40189 CATGGCGGCTGATGCGGGGACGCGGACTGAAGTCTGAGACACCGCTCGCGCGAGA 40130
Qy 1004 GCTC-----GACGCGCGGATATCGGACGAGCGGAGCGGAGCGGCGGAGGA 1048
Db 40129 ACTCCAGAGGACAAAGGAGCGGCAAGGAGCTCGAGGCTCGGAGATCTACCTGCGAAA 40070
Qy 1049 TGGCTACAGCTGCGCTCGCGCTCGCAGGTCTCGGGGCGGCTTCGACAGCTCGCATG 1108
Db 40069 GGGCTACTCGTGGCGGCACTCCCGGAGTCTCGGGGCGGCTGGCGGACACCTTGTACCA 40010
Qy 1109 GCATGACCGGCTGTGACGATCGAGCTGAACCGGCTGACGCAATCCGCTGTTTCGCC 1168
Db 40009 CGCGGSCACAGCTGCGCATCGAGCTCACTCGGCCAAAGCAACCGCTCTT----- 39956
Qy 1169 CGATGGCAGCTGCCCGCTGACGCGGGGCAATTTTCATGGCGCAGCATGTGGCGCTGAC 1228
Db 39955 CTTGAGGGCAAGGAGATCTTCCACGGGGGAACTTCCAGCGTCAAGCGGCTTCG 39896
Qy 1229 GTCCGATGCGCTCGCAGCGCGCTCACCGTTCTGGCGGCGCTTGGAGCGCCAGATGC 1288
Db 39895 GATGGACTTCGTGACCATCGCGCTCACCCAGCTCGCGCTCTGGCGGAGCGGAGATCAA 39836
Qy 1289 ACGTCTGACAGATGAAGGCTGAACCGTGGCTGCGCCCTTCTCCACCGGGCGCCCGC 1348
Db 39835 CCGGCTCTGAAACCGGACCTCAGCTAGCGCTCCCGAGTTCCTGTCTCGGGGAGCC 39776
Qy 1349 CGGTTGAAATTCGGCTTTCATGGCGCAGAGGTGAACCGGACCGCGCTCTCTGGCGGAGAT 1408
Db 39775 GGGCTGCAAGCGGATTCGCGCGGCGCCAGTACCGGCGCACCGCACTGGTGGCGGAGAA 39716
Qy 1409 GCGAGCACCGGACCTCGCTCGATTCATTCGATTCACGAAACCGCGCCCAATCGAGATGT 1468
Db 39715 CCGGACCATCGCGCGCGGCGGACCCAGAGGCTCCCGTCCAAACCGCGCACCAACAGGACGT 39656
Qy 1469 GGTCTCGCTTGGGACCATCGCGCGCGCTCTGCGCGGAGAGATCGACCGTTGGGCGGA 1528
Db 39655 GGTGAGCATGGGCTGTATCTCGGCGCGCGGAGCGCGGCTGCTGTGCGAACAAACAA 39596
Qy 1529 GATCTTGGCATCTCGCTCTCTGTCTGACAGCTGCGGAGCTGCGGCTGCGGCGAGCGG 1588
Db 39595 GATCTCGCG-----GTGGAGTACCTGGCGCGCGCGCGGAGGCTGACATCTCGCGCG 39542
Qy 1589 CCTAGACGGGCTGTCTCCCGCGGGGAAAGCTGGTGTGACGCGCTCGCGGAGCAGTTCCC 1648

Db 39541 GTTCGACGGCTTGAGCCCGGCGGAAGCCACGCTACGAAGCGGTGCGCGCTGTTCC 39482
Qy 1649 GCCTCTTGAGACGAGACCGGCCCTTGGGACAGAGAAATTGCCCGCTTGC 1696
Db 39481 GACCTGGGCGTCCACCGGTACATGTCGGCGACGACATGAGTGGTCCG 39434
RESULT 7
US-10-292-198-1/c
; Sequence 1, Application US/10292198
; Publication No. US20030157654A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: LIU, Wen
; TITLE OF INVENTION: BIOSYNTHESIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE
; FILE OF INVENTION: PATHWAY
; FILE REFERENCE: 054030-0007
; CURRENT APPLICATION NUMBER: US/10/292,198
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: US 10/159,257
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 09/478,188
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: US 60/115,434
; PRIOR FILING DATE: 1999-01-06
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 63158
; TYPE: DNA
; ORGANISM: Streptomyces globisporus
; US-10-292-198-1

Query Match 13.1%; Score 242.4; DB 6; Length 63158;
Best Local Similarity 50.3%; Pred. No. 1.1e-53;
Matches 749; Conservative 0; Mismatches 706; Indels 33; Gaps 5;
Qy 227 CGTGGCAGCGCGCGCGCGGATTTGCTTGGCCCTTCGCGCGCGGACCGGTGCGGTGC 286
Db 40901 CGTGGCAGGAAACGCGCGGCTCGACGTACCGCGCAATCCATCGGAAGGCCAGAA 40842
Qy 287 GTCCGAAGCGCGCTCGCGCTGTCTATCCGCGAGCGCGCATGTCTACCGACTGACAC 346
Db 40841 GAGCGCGGAGATCTTCGAAGGGATCGCGCAACAGAACATCCCCATCTACGGGGTAC 40782
Qy 347 CGGCTTCGCTCCCTTCGGAACCGCTGTCTCAGGTGAGAACTCGGAACGCTGCAGGC 406
Db 40781 CGGTCAGCGGAGATGATCTACATGAGGTGCAACAGTCGAGAAAGTCGACTGACAG 40722
Qy 407 CAATCTTTGTCATCATCTGGCCAGCGCGGTGGGACCGGTGCTTGAATGGACGACGGCG 466
Db 40721 CAATCTCGTCCGTAGCCACAGCGCGGAGTCGGTCCGCTGTTCCGCGAGGAGGCGCG 40662
Qy 467 GCGCATGTTCTGGCGGTCTGGTGTGATGCTCAGGAGGACCTCCGCTGCGACGAGGG 526
Db 40661 GGGGATGCTGCGCGCGCGGCTGAACACCTCGCAAGGGCCACTCCCGGCTGCGCCCAT 40602
Qy 527 GACCATCGCTCGGCTGATTCGACCTGTCAATTCGAGCTCGCTCCGCGCGCTTCCAGCG 586
Db 40601 CATCTCGAACGCTCTCGGCACTTGAACGAGGGCATCACCCGCGCATACCCGAGAT 40542
Qy 587 CGGCAACGTTGGCGCGCTGCGGTGACCTGACACCGCTTTCGCGATATGGTGTCTTCCCTCA 646
Db 40541 CGGTCATCTCGGCGGAGCGCGGACCTGGCTCCCTCTCCCATCGTCGCGAGCACCTCAT 40482
Qy 647 GGGCGGGGAGACTTCTCTGACCGGAGCGGAGCGGAGCTTGAACGCGCGAGAGGGCTCCG 706
Db 40481 CGGAGAGGGCTAGCTCTCTG---CGCGACGAGCGCGGTCGAGACCGCCAGGTGCTGGC 40425
Qy 707 GCGCGGACGCTGCAACCGCTCGATCTCTCCCATCGGATGCACTGGGCTGGTCAACGG 766
Db 40424 CGAGCGGGGATCGAGCCGCTCGAACTGCGCTTCAAGGAGGGCGCTCGCACTGATCAACGG 40365

QY 767 GACCTCGGCATGACCGGAGTCGGCTGTGTGAATGTCTACGCTGCGGCATATCTCGCAA 826
Db |||||
40364 CACGTCGGGATGACCGGTCTGGGCTCCCTGTGTCGTGCGAGCGGCCCTTGAGCAGGCCCA 40305
QY 827 CTGGCGGCTGGCTTACCGCCCTGTCTCGGAAATCTTGAAGCGCGGACCGAGCATG 886
Db |||||
40304 GAGGCGGAGATCGTACCGCTCTGTCTATCGAGGCGGTACCGGGATCGACACGCCCCCTT 40245
QY 887 GGCCTGGGCACCTGTCGCACT---GGGGCGCATCCCGACAGAAAGCGCCGACCGAG 943
Db |||||
40244 CCTCGCGAGGGSCACGACATAGCCCGCGCACGAGGCCAGATCGACACCGCCGCCAA 40185
QY 944 GCTGCGGCCCGCGTGGACCGGAGCGCGGGTGTCTCCGACAGTCAATTTGCGAGCGAG 1003
Db |||||
40184 CATGCGGGCCTGATGCGGGGACGCGACTGACGCTCGAGCACGCCGACCTGCGCCGAGA 40125
QY 1004 GCTC-----GACCGCGCATATCGGGACGGAGCGGAGCGGGGAGCA 1048
Db |||||
40124 ACTCCAGAGGACAAGAGGCGCGGCAAGGACGTCCAGCGCTCGGAGATCTACTCTGCAGAA 40065
QY 1049 TGCCTACAGCTCGCTGCGCTCCGCAAGTTCCTCGGGGCGGCTTCGACACGCTCGCATG 1108
Db |||||
40064 GGCTACTCGCTGCGGGCATCCCCAGGTCTGCGGGCGGTGCGGACACCTTGTACCA 40005
QY 1109 GCATGACCGGGTCTGACGATCGAGTGAAACGCGGTGACGCAATCCGGTFTTCCGCC 1168
Db |||||
40004 CGCGCGGCAAGACTCGCATCGAGCTCAACTCGGCCAACGACAAACCCGCTCTT----- 39951
QY 1169 CGATGACAGGTGCCCGCTTGACGGGGCATTTTCATGGGCGCAGCATGTGGCGTGAC 1228
Db |||||
39950 CTTGAGGGCAAGAGATCTTCCACGGGGCGAACTTCCACGCTCAGCCGATCGCGTTCG 39891
QY 1229 GTCCGATGCGCTCGCCACGCGCTCACCGTTCGTGGCGGCGCTTGGAGCGCCAGATTCG 1288
Db |||||
39890 GATGGAATTCGTGACCATCGCGTCAACCGAGTCTGCGGCGGCGGAGATCAA 39831
QY 1289 AGCTCTGACAGATGAAGGTGAACCGTGGGTGCCCCCTTCTTCCACCGGGGCCCCGC 1348
Db |||||
39830 CGGGTCTTGAACCGGCACCTCAGCTACGCGCTCCCGGAGTTCCTGCTCCGGGAGCC 39771
QY 1349 CGGTTGAATTCGGCTTCAATGGGCGCAGAGTGAGCGGCGACGCGCTCTTGGCCGAGAT 1408
Db |||||
39770 GGGGCTGACAGCGGATTCGCGGGCGCCAGTACCAGCGGCGACCGCACTGGTGGCGAGAA 39711
QY 1409 GCGAGCAGCGGACCTGCTCGATCCATTCGATCTCCAGAAACCGCGCCCAATCAGGATGT 1468
Db |||||
39710 CCGGACGATCGGCCCGCCAGCACCCAGAGCGTCCCGTCCACGCGGACAAACAGACGT 39651
QY 1469 GGTCTCGCTTGGACCATCGCGCGCGCTCTGCGCGGAGAGATCGACCGTTGGCGGA 1528
Db |||||
39650 GGTGAGCATGGGCTGATCTCGGCCCGCAACCGCCCGCGGCTCTGTTCGAAACAAACAA 39591
QY 1529 GATCTTGGATCCTCGCTCTGTCTTCCACAACTGCGGAGCTGCGTGGCGGACGG 1588
Db |||||
39590 GATCTCTCGG-----GTGGAGTACTGTCGCGCGCGCCAGGCGGTGCAATCTCGCGCG 39537
QY 1589 CTTAGACGGGTGTCTCCCGCGGGGAAGAGTGTGTGTCAGGCGCTTCCGCGAGCAGTTCC 1648
Db |||||
39536 GTTCGAGGCTTGAGCCCGCGGCGGAGGCCAGTACGAGCGGTGCGCGGCTGTTCC 39477
QY 1649 GCGCTTGGACGAGCCGCCCTTGGACAGGAAATTCGCGCGCTTGC 1696
Db |||||
39476 GACGCTGGGCTCGACCGGTATATGSCCGACGACATCGAGCTGGTTCG 39429

RESULT 8

US-10-369-493-31670
; Sequence 31670, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31670
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31670

Query Match 12.9%; Score 238.4; DB 6; Length 1497;
Best Local Similarity 50.8%; Pred. No. 1.3e-52;
Matches 739; Conservative 0; Mismatches 666; Indels 51; Gaps 5;

QY 199 GCCACATCGATCTGGACACAGGCCCATCCGTCGGCGAGCGCGCGCGCGGATTTGTCCTTG 258
Db |||||
14 GCCAAGCCACGCTCGCGAGCTCGAAGCATCTGGCGGAGGGCGCGCGCGCTTGG 73
QY 259 CCCCTCCGCGCGCGACCGGTCCTGCTGCCAAGCGCGCTCGCGCTGTGATCCCG 318
Db |||||
74 CCCCAGAGCGCGCCCGCGCTCGAGGCGCGCGCGCGCTGGCCGAGGCGCGCGCG 133
QY 319 AGCGCGGCATCTTACGACATGACACCGGCTTGGTCCCCTTGGGAACCGCGCTGATCT 378
Db |||||
134 GCACGCGCGCGCTCTATGCGCTGAAACAGGGCTTCGGCAAGCTCGAGACCTCAAGATCG 193
QY 379 CAGGTGAGAAATCTCCAAACGCTCGAGGCCAATTTGTTCATCATCTGGCGAGCGGCTGG 438
Db |||||
194 CTCGCGCGATAGCGGCGCACTGCAACGCACTGATCTGTGCACTCTGCGGCGTGG 253
QY 439 GACCGGTGCTTGAACGACGCGCGCGCATGTTCTGCGCGGTCTGGTGTGATCG 498
Db |||||
254 GCGAGCTATGCCCCCGTCCACGCGCGGCTGATGATTCGCTGAAGCTCTGTGCTCG 313
QY 499 CTCAGGAGCTTCGCGTCCAGGAGGAGACCATCGCTCGCGTGAATCGACCTGCTCAAT 558
Db |||||
314 GCGCGCGCGCTTCGCGCGGTGCGTGGAGATCGTGGCGCTGCTCGAAGGCAATGCTGGCG 373
QY 559 CCGAGCTCGCTCCGCGCGTTCACAGCGCGCACGCTGGCGCGCTCGGCTGACCTGACAC 618
Db |||||
374 CCGGCGTCAACGCGGTGATCCCGCGCAGGGTCTGCTCGCGCGAGCGCGATCTGGCAC 433
QY 619 CGCTTGGCATATGTGCTCTGCTTCCAGGCGCGGAGACTTCTTGAACCGGAGCGGA 678
Db |||||
434 CCCTCGCCCATATGCGCGGCTCATGATCGCGAGGCGAGGCCAGGTC---GGCGGCA 490
QY 679 CCGGCTTGAAGCGCGCAGAGGCTCCCGCGCGGACGCTGCAACCGCTCGATCTCTCC 738
Db |||||
491 GCGCGCTGCGCGGTGCGCGGCGCTGCGCGAGCGCGCTTTCGCCCGGTGGCGCTCGGAC 550
QY 739 ATCGCATGTCATCGCGCTGTGTCACAGGACCTCCGCCATGACCGGAGATCGCGCTGGTGA 798
Db |||||
551 CCAAGGAAGGCTCGCCCTCATCAACGCGACGCAATTTCTCGACCGCTATGCTCCCGCG 610
QY 799 ATGCTCAGCGTCCCGCATCTCGGCAACTGGCGGCTGGCTGAGCGGCTTGTTCGG 858
Db |||||
611 GCCTCTTCAGGGCTGGCGCGCTCAGGCGGCGCTTGTGATCTCGGCGCTCTCCACCG 670
QY 859 AATGCTGAGAGCGGACCGAGGATGGGCGCGGCACTGTCGACCTGCGCGCGCATC 918
Db |||||
671 ATGCGATCATGGGCTCGACTGCGCGCTCCCGCCGAGATCCACCGCTGCGCGGCAATG 730
QY 919 CCGGACAGAGGACCGCGCAGCGAGCTGCGCGCGCGCTGAGCGGCGCGGCTGG 978
Db |||||
731 CCGGCGCATGAGGCGGCGCGCACCATGCGCGGCTCTGCTGAAAGCTCGGCCATCCGCG 790
QY 979 TCCGCGCATATTCGCGAGCGGCTCGACCGCGCGGATATCGGAGCGGAGCGGAGG 1038


```
Db 791 AGAGCCACCGTGTAGGCGGACC-----AGC 814
Qy 1039 CGGGGAGGATGCTACAGCCTGGCTGCGCTCGCGAGGTTCTCGGGCGGGCTTTCAGACA 1098
Db 815 GGGTGCAGGACCCCTACTGTCATCGCTGCGAGCGCAGGTGACGGGGCGCGCATGTGATG 874
Qy 1099 CGCTCGCATGTCATGACCGGTGCTGACGATCGAGCTGAAACGCGGTGACCGACAAATCCGG 1158
Db 875 TGCTGCGATGCGCGCGGACGCTGGCCACCGAGGCCAATGCCCGCACCGACAATCCGC 934
Qy 1159 TGTTCGCGCCGATGCGAGCGTGGCCCTCTGACAGGGGGCAATTTTCATGCGGCCAGCATG 1218
Db 935 TCGTGTCTCGGA-----CGGGCGCATGCTCTCGGAGGCCAATTTTCATGCGGAGCCCG 988
Qy 1219 TGGCGCTGACGTCCGATGCGCTGCCACGCGCGTACCGTTCTTGGCGGGCTTTCGAGAC 1278
Db 989 TGGGCTTCGCGCGCGACATGATCGCTGGCGCTCTCCGAGATCGGGCCATCGCGCAGC 1048
Qy 1279 GCCAGATTGACGCTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCTTCTCTCCACC 1338
Db 1049 GCGCGTGGCGCTGATGGTGATCCGACGCTCTCTTCGACCTTCGCGCTTCTCTCACCC 1108
Qy 1339 GGGGCCCGCGCGGTGAATTCGCGCTTCATGGCGCACAGGTGACCGCGACCGCGCTCC 1398
Db 1109 CCGAGC---CGGGCTGAATTCGCGGCTGATGTCGCGAGTGACGACGGCCGCGCTCA 1165
Qy 1399 TGGCCGAGATGCGAGCCACCGGACCTGCTGATFCCATTCGATCTCACGAACCGCGCCA 1458
Db 1166 TGACGAGAAACAAGACATGGCGCCGCCACCGTCAACGACAGCACGCCACCTCGGCCA 1225
Qy 1459 ATCAG---GATGGTCTCGTTGGGACCATCGCGCGCGGCTCTGCGCGAGAGATCG 1515
Db 1226 ATCAGGAAGATCATGATGATGCGGCCCATGCGCGCGCGAGGCTCGGCGCGATGCTG 1285
Qy 1516 ACCGTTGGCGGAGATCTTGGATCTCGTCTGCTCTGTTGACAAAGCTGCGGAGCTGC 1575
Db 1286 AGAACCTCGCGGTGATCTCGGACCGAGGAGATCTCGCGCGCCAGGAGGTGAGTTCC 1345
Qy 1576 GCTCGCGAGCGGCTAGAGCGGTGTCTCCCGGGGAAAGACTGCTGACGCGCCCTGC 1635
Db 1346 GCGCGCCCTCGCCACCTCGCGCCGCTCGCGCGCGTGGCGCGGTGGCGCGGAGG 1405
Qy 1636 GCGAGAGTTCCCGCC 1651
Db 1406 TGC CGCGATCGCGGC 1421
```

RESULT 9

```
US-10-156-761-3312
; Sequence 3312, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3312
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1542)
US-10-156-761-3312
```

```
Query Match 12.6%; Score 232.8; DB 6; Length 1542;
Best Local Similarity 50.0%; Pred. No. 3.8e-51;
Matches 730; Conservative 0; Mismatches 682; Indels 48; Gaps 4;

Qy 215 CAGAGCCCATGCGGTGCGAGCGCGCGCGGATGTCTTCCCTCCGCGCGCGA 274
Db 48 CGAGTTCTCGCCCTGCGGTGCGCGCGCGCGATCGAGCTCTCCGGGAGGGGTGAC 107
Qy 275 CCGGTGCGGTGCGGTCCGAAGCGCGGCTCGGGCTGTCTATCCGCGAGGCGCCATGTCTA 334
Db 108 CGCCCTCGCGCGCGCGCGCATGCTGGACGCGCTGGCGGCCAAGCGGAGCCCGTGTGTA 167
Qy 335 CGGACTGACAAACCGGCTTCGCTCCCTTGGGAAACCGCTGATCTCAGGTGAGAAATGTCCG 394
Db 168 CGGGGTGAGCACCGGCTTCGGGGCCCTGGCGACCCGGCACATCAGCCAGGAGCTCGGGC 227
Qy 395 AAGCTGACAGGCCAATCTGTTCATCATCTGGCGAGCGGCTGGGACCGGTGTGACTG 454
Db 228 GCAGCTCCAGCGCAACATCTGTCGCTCGCACGCGCGCGCATGGTCCGAGGGTGGAGCG 287
Qy 455 GACGACGGCGCGGCCATGCTTCTGGCGCGTCTGGTGTGATCGCTCAGGAGGCTCCGG 514
Db 288 GGAGGTGCTACGGCGCTGATGTTCTTGGGACTGAAACCGCTCTGCTCGGGCCACACCGG 347
Qy 515 TGCAGCGAGGGGACCATCGCTCGCTGTATCGACCTGCTCAATTCGAGCTCGCTCCGGC 574
Db 348 CGTAGCGCCCGAAGTCCGCGAGAGATGGCGGACATCTCAACCGCGGATCACCCGCT 407
Qy 575 CGTTTCCAGCGCGGACCGGTGGCGGCTCGGGTACCTGACACCGCTTGGCGATATGCT 634
Db 408 CGTGCAAGTAGTACGGCTCGCTCGGCTGCTCGGCGACCTCGCCCGCTCTCCCACTGCGC 467
Qy 635 GCTCTGCTCCAGGCGCGGGAGACTTCTTGAGACCGGAGACGGGCGGTGTGAGGGCGC 694
Db 468 CCTGACGCTGATGGCGAGGGCGACGCGGAGGGCCGCGACCGCTCGCTCGCGCGCG 527
Qy 695 AGAAGGCTCCGGCGGAGCGGCTGCAACCGCTCGATCTCTCCCATCGCGATGACTGGC 754
Db 528 TGACCTCTCGCGAGCACGGCTACGCGCGGTGAGCTGCGGAGAGGAGGCGCTTGGC 587
Qy 755 GCTGCTCAACGCGGACCTTCGCGCATGACCGGGATGCGCTGCTGTAATGCTCACGCTGCGC 814
Db 588 CCTCTCAACGCGCACCGACGCGCATGCTCGGCATGCTGCTCATGGCCCTCGCGACCTGGA 647
Qy 815 CCATCTCGGCACTGGGCGGTGGGTGACGGGCTCTGTCGGGAATGTCTGAGAGGCGC 874
Db 648 CGCGCTCTACAAGTCTGCGGACATCACGGCGGCGCTCTCGCTCGAGGCACTCTCTCGGCAC 707
Qy 875 GACCGAGGATGGGCGCGGCACTGTCCGACCTGCGGCGCATCCGCGAGAGGAGCGC 934
Db 708 CGAAGAGTCTCGCCCCGAACTGACACGCTATCGGCGCGACCCCGCTCAGGGCGGCTC 767
Qy 935 CGCAGGAGGCTCGGGCGCGCGGTGAGACGCGAGCGCGCGGTGCTCGGCGACGCTATTGC 994
Db 768 GCGCGCAACATGCTGGCGGTACTGGCGGGTTCGAGC-----805
Qy 995 CGAGCGAGGCTCGAGCGCGCGGATATCGGGAACGAGCGCGGAGCGGGGAGGATGCTA 1054
Db 806 -----TCACCGGCCACCAACAGGACGACGCGCGCGCTCCAGGACGCGTA 851
Qy 1055 CAGCCTGCGCTCGGCTCGGCAAGTTCTCGGGGCGGGTTCGACACGCTCGCATGCGATGA 1114
Db 852 CTCGCTGCTGCGCGCGCGAGTTCGCGCGCGCGCGCGGACACCTTCGCCACGCGCG 911
Qy 1115 CCGGGTGTGACGATCGAGCTGAACGCGGTGACGCAATCCGGTGTTCGCGCGGATGG 1174
Db 912 CCTGTCGCGCGGAGGAGCTGGGCTCGGCGGTGGACAAACCGGTGTGCTTCCCGA---968
```

1175 CAGCGTCCCGCCCTGACGGGGGCAATTTCATGGGCCACGATGTGGCGTACGTCCGA 1234
Db ---CGGGCGCGTGGAGTCCAAACGGCAATTCACGGTGCCTCGGTACGTCCTCGA 1025
1235 TGCGCTCGCCACGGCGCTCACCGCTTCCTGGGGGCGCTTGGCGAGCGCCAGATTGCACTCT 1294
Db CTTCTCTCGGATCGCGCGGCCGACTTCGGGTCCATCGCGAGCGCGCACCGACCGGCT 1085
1295 GACAGATGAAGGCTGAACCGGTGGGCTGCCCGCCCTTCCTCCACCGGGGGCCCGCGGGTT 1354
Db TCTCGACAAGAACCGCTCGACGGGCTGCGCGGTTCTCT---CGCGACGACGCGGCGT 1142
1355 GAATTCGGGCTTCATGGGCGCAAGGTGACGGCGACCGCGCTCCTGGCGAGAT---GGG 1411
Db CGACTCCGGCTGATGATCGCCCACTACACGAGCGGGCCCTGGTCAGCGAGATGAAGCG 1202
1412 AGCCACGGGACCTGCTCGATCCATTGATCTCCAGAAAGCGCCCAATCAGGATGTGT 1471
Db GCTCGGCTCCCGCGCTCGCGACTCGATCCCGTCTCGGCGATCGAGAGGACCACT 1262
1472 CTCGCTTGGGACCATCGCGCGCGCTCTGCCCGAGAAAGATCGAACCGTTGGGCGGAGAT 1531
Db CTCGATGGGTGTGCGCGCGCGCAAGCTCTCGTACGGCGATCGACAACTCAGCGCAT 1322
1532 CTTTGGATCTCGCTCTGTCTTGCAAGCTGCGGAGCTGCGTGGCGAGCGGCT 1591
Db CFTGGCGATCGAGCTTACGCCGCCACCGCGCATCGAACTCCGCGAGGCTCTGACCCC 1382
1592 AGACGGGGTCTCCCGCGGGAAGCTGTGTGAGGCCCTGCGGAGCAGTTCGCGCC 1651
Db CGCCCCGGGTGCGAGCGGTATCATCGGCCGTCCGAAAGGCGGCGTGAAGGCCCAAG 1442
1652 GCTTGAGACGACCGGCGCC 1671
Db ACCGACCGCTTCTGGCCC 1462

RESULT 10

US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Query Match 12.6%; Score 232.8; DB 6; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 3.6e-51;
Matches 730; Conservative 0; Mismatches 582; Indels 48; Gaps 4;
QY 215 CCAGGCCCATGCGTGGCGGCGGCGGCGGATTTGCTTGCCCTCCGCGCGCGA 274

Db 4139678 CGAAGTTCTCGCGTGGCGGTGGCGGCCCGCGATCGAGCTCTCCGGGAGCGGTGAC 4139619
QY 275 CGGTGTCCTGGCTCCGAAGCGCGGTGCGCGCTGTCATCCGCGAGGCGCGCATGTCTA 334
Db 4139618 CGCCCTCGCGGGCCCGGGGATCGTGAACGCGCTGGCGGCCAAGCCGAGGCCGTGTA 4139559
QY 335 CGGACTGACAACCGGCTTCGGTCCCTTCGAAACCGCTGATCTCAGGTGAAGATTCGG 394
Db 4139558 CGGGGTGAGCACCGGCTTCGGGGCCCTGGCGACCCGGCACATCAGCCAGAGCTCGGGC 4139499
QY 395 AAGCTGAGGCGCAATCTTGTTCATCATCTGGGCAGCGGCTGGGACCGGTGTTGACTG 454
Db 4139498 GCAGCTCCAGCGCAACATGTCCTCGCTCGACCGCGCGCATGGTCCGAGGTTGAGCG 4139439
QY 455 GACGACGGCGCGCGCATGTTCTGGCGGTCTGGTGTGATCGCTCAGGAGCCCTCCGG 514
Db 4139438 GAGGTGTAACGGCGCTCATGTTCTGCGACTGAAAAACGCTGTGTCGGGCCACACCGG 4139379
QY 515 TGCCAGCGAGGGGACCATCGCTCGCTCGCTCAATTCGAGCTCGCTCCGGC 574
Db 4139378 CGTACGGCCGAAGTCCGCGACAGATGCGCGACATCTCTCAACGCCGGGATCACCCGGT 4139319
QY 575 CGTTCCACCGCGGACCGGTGGGCGCTGGGTGACTGACACCGCTTGCAGATATGTT 634
Db 4139318 CGTGAACGAGTACGGCTCGCTCGGCTGCTCCGGCGACCTCGCCCGCTCTCCACCTGCGC 4139259
QY 635 GCTCTGCTCCAGGGCGGGGAGACTTCTGACACGGGACGGGACCGCGCTTACGCGCGC 694
Db 4139258 CTTGAGCTGTATGGGAGGGGCGACCGGAGGGCCCGGACGCCGCTCGTCCGCGCCG 4139199
QY 695 AGAAGGGCTCCGCGCGGACGGCTGCAACCGCTCGATCTCTCCCATCGGATGCACTGGC 754
Db 4139198 TGACCTCTCTCGCGAGCAGCGCATCACGCCGCTCGAGCTGCGGAGAGGAGCGCTGGC 4139139
QY 755 GCTGTGTAACGGGACCTCGGCCATGACCGGATGCGGCTGGTGAATGCTCAAGCTGCGC 814
Db 4139138 CCTCTCTAACCGCACCGACGGCATGCTCGGCGATGCTGGTCTATGGGCCCTCGCGACCTGGA 4139079
QY 815 CCATCTCGGCAACTGGGCGGTGGCGTTGACGCGCTCTGCGGATGTCGTAGAGGCGCG 874
Db 4139078 CGGGCTCTCAAGTCTGCGGACATCACGGCGGCCCTCTCCCTCGAGGCACTCTCTGGCAC 4139019
QY 875 GACGAGGCAATGGCGCGGCACTGTCCGACCTGCGGCGCATCCCGGACAGAAAGACGC 934
Db 4139018 CGACAAGGTGCTCGCCCCGAACTGCACGCCATCCGCCCGCACCCCGTCAGGGCGCTC 4138959
QY 935 CGCAGGAGGCTGCGCGCCCGCGGTGGAACGCGAGCGCGGGTGGTCCGGCACGTCATTGC 994
Db 4138958 GGCGCGCAACATGCTGGCGGTACTTGGCGGTTCCGAGC----- 4138921
QY 995 CGAGCGGAGGCTGAGCGCGCGGATATCGGACGGAGCGGAGGCGGCGAGTGCCTTA 1054
Db 4138920 -----TCACCGGCCACACACGAGGACGCGCGCGGTCTCAGAGCGCTA 4138875
QY 1055 CAGCTTGCCTGCTCGCTCCGACAGTTCTTCGGGGCGGGCTTCGACACGCTCGCATGGA 1114
Db 4138874 CTCGCTGCGCTCGCGCCCGCAGTCCGCGCGCGCGGACACCTTCGCCACGCCCG 4138815
QY 1115 CCGGGTGTGACGATCGAGCTGAACCGGTGACCGCAATTCGGTGTGTTTCGCGCCGATGG 1174
Db 4138814 CCTCGTCCGCGAGCGGAGCTGGCGTGGCGGTGGAACAACCGGTGTGTCCTCCGA--- 4138758
QY 1175 CAGCGTGGCCCGCTCGACCGGGGCAATTTTCATGGGCCAGCATGTGGCGCTGACGTCCTGA 1234
Db 4138757 ---CGGCGCGCTCGAGTCCACCGGCAATTTTCAGGTGCGCGCTGCTGCTCTCGA 4138701
QY 1235 TCGCGCTCGCCACGGCGTCAACCGTTCTGGCGGGCCCTTGGGAGCGCGCAGATTGCACTCT 1294
Db 4138700 CTTCTCGGATCGCGCGGCGGCGGCTCGGGTCCATCGCGGAGCGCGCACCGCGGCT 4138641
QY 1295 GACAGATGAAGGCTGAACCGGTGGGCTGCCCGCTCTCTCCACCGGGGGCGCGCGGTT 1354

Db	4138640	TCTCGACAAAGAACCGCTTCGACAGGGCTGCGCGGTTCCT---	CGCGACAGACGGGGCGT	4138580
Qy	1355	GAATTCCGGCTTATGCGGCACACAGGTGACGGGACGCGCTCTCTGGCCGAGAT---	GCG	1411
Db	4138583	CGACTCCGGCTGATGATCGCCAGTACACGACGGCGCCCTGCTCAGCGAGATGAGCG	4138524	
Qy	1412	AGCACGGGACCTGCCTTCGATTCATTTCGATCTCCACGACGCCGCCAATCAGGATGTGGT	1471	
Db	4138523	GCTCGCGGTCCCGCGCTCCGCGACTCGATCCCGTCTTCGCGCATGCAGGAGGACCACGT	4138464	
Qy	1472	CTCGCTTGGGACCACTCGCGCGGCGCTCTCGCGCGAGAAATCGACCGTTGGCGGGAGAT	1531	
Db	4138463	CTCGATGGGTGTGTGCGCGCCGCCAAGCTCCGTAAGCGCATCGACAACTCAGCGGCAT	4138404	
Qy	1532	CCTTGCATCCTCGCTCTCTGTCTTGACAAGCTGCGGAGCTGCGCTGCGGACAGCGCCT	1591	
Db	4138403	CGTGGCGATCGAGCTTACGCGCGCACCCGCGGCATCGAACTCGCGAGGGTCTGACCCC	4138344	
Qy	1592	AGACGGGGTGTCTCCGCGGGGAGAGCTGGTGTCAGGCCCTGCGCGAGCAGTTCGCGCC	1651	
Db	4138343	CGCCCCGGCTGCGAGCGCGTCTACAGCGCCCTCGCAAGCGGGCGTTCGAAGGCCCAGG	4138284	
Qy	1652	GCTTGAGACGACCGGCCCC	1671	
Db	4138283	ACCGGACCGCTTCTTGCCCC	4138264	

RESULT 11

```

US-10-282-122A-13842
; Sequence 13842, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: wang, Liangsu
; APPLICANT: zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13842
; LENGTH: 1521

```

Db 965 TGTGGGCGGCAATTTCCACGCGAGCCGCTCGTTCGCGGCGGCAATCTCGCATCG 1024
Qy 1249 CCGTACCGTTCTGGCGGGCTTTCGAGCGCGCAGATTCAGCTCTGACAGATGAAGGC 1308
Db 1025 CCGCGGCCAGATTCGGCGGCTCGCGAGCGCGCATCGCGTGTGATGACGCGACGC 1084
Qy 1309 TGAACCGTGGGCTGCGCCCTTCTCCACCGGGGCGCGCGGGTTGAATTCGGGCTTCA 1368
Db 1085 TCTCCGCGCTGCCGCC-----TTTCTCGTGAAGGACGGCGTGAATCGGGCTTCA 1138
Qy 1369 TGGGCGCACAGGTGACGCGGACCGCGCTCTGCGCGAGATGCGAGC---CAGCGGACCTG 1425
Db 1139 TGATCGCGACGTGACGCGCGCGCTCGCGTCGGAAACAAAGACGCTCGCGCATCCCG 1198
Qy 1426 CCTCGATCATTGATCTCCACCAACGCCCAATCAGATGTGCTCTGGGACCA 1485
Db 1199 CGTCGGTCGATTCGCTGCGCAGCTCGCGGAACAGGAAACACAGTGTGATGGGACGT 1258
Qy 1486 TGCCCGCGCGCTCTCGCGGAGATCGACCGTTGGGCGGAGATCTTTCGATCCTCG 1545
Db 1259 TCGCGCGCGCAAGCTCACGGACATCGGAGAACGTTCGGAACATCTTCGATCGAGC 1318
Qy 1546 CTCTCTGTCTTGACAACTGCGGAGTGTGCTGCGGAGCGGCTTAGACGGGCTGTCT 1605
Db 1319 TGCTCGCGCGCGCAAGCGTCTGACCTGC-----GCGCGCGCACGCAACGAGCC 1369
Qy 1606 CCGCGGGGAGAGCTGTGCGAGCCCTCGCGAGCAGTTCCGCGCTTGAGACGGACC 1665
Db 1370 CGCGCGTGCAGCAGCGATGAAGACGATTTCGCGCGGACGTCGCGCCTACGATCTCGACC 1429
Qy 1666 GGCCCTGGGACAGGAAATTCGCGCGCTGTGCTACGCACTT 1705
Db 1430 ACTACTTCGGCCCGACATCGCGTGGTTCGGCGGCGCGT 1469

RESULT 12

US-10-369-493-42898
; Sequence 42898, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: PLANTS WITH IMPROVED PROPERTIES
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42898
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-10-369-493-42898

Query Match 11.6%; Score 215.4; DB 6; Length 1503;
Best Local Similarity 50.0%; Pred. No. 1.4e-46;
Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
Qy 203 CATCGATCTGGACCGCCATCCGTGGGAGCGCGCGCGGCGGATGTCTTCCGCC 262
Db 21 CCTGAAGCTCGAGAAATCTCCAGTGGCTCGCAACGAGGCCACCGTGGAGCTGTCCG 80
Qy 263 TCGCGCGCGACCGGTGCGTCCGAAGCGCGGCTCGGGCTGTTCATCGCGAGGC 322
Db 81 CGAGGCGGCACCGGTGGGCGCTCGCGCGCTGTGGACCGGTGTCCGCCCGAGA 140
Qy 323 GCGCCATGTCTACGGACTGACAAACCGGCTTCGGTCCCTTTCGGAACCGCGCTGATCTCAGG 382

Db 141 CACGCCCGCTACGGCATCAACACCGGCTTTGGCAGTGTGGCGGAGGTCCGATCGACAA 200
Qy 383 TGAGATGTCCGNAACGCTCAGCGCCAAATCTTGTCCATCATCTGGCCAGCGGTGGGACC 442
Db 201 GAAGGACCTTGCGBACCTCCAGCGCAACCTCATCTCTCCACGCGTGTGGCGTCGGCAC 260
Qy 443 GGTGCTTGACTGACACGCGCGCGCCCATGGTTCTTGGCGCGTCTGGTGTCCGATCCTCA 502
Db 261 GCGCTTCCCTTCCGGAAGCGGGGCGCTCTGTCTCGCTCGCTGCAACGTGTCTGCCAA 320
Qy 503 GGGAGCTCCGTTGCGAGGAGGACCATCGCTCGCTCGCTGATGACGCTCTCAATTCGGA 562
Db 321 GGGTACTCCGCGATCCGATGAGACGCTGCGCTTGGCGCTGGAATGCTGAACCGGGA 380
Qy 563 GCTCGCTCCGCGCGTTCGCCAGCGCGGCGACGCTGGGCGCGTGGGAGCTGACCTGACACCGCT 622
Db 381 CGTGTGCGCGTGGTCCCGAGCGGGGCGCTCGGGGATCTCGCCCGCT 440
Qy 623 TCGCATATGTGTCTCTGCTCCAGGGCGGGGAGACTTCTTGACCGGGAACGGGACGCG 682
Db 441 GGGGACCTTGGCTGTCTTTCATCGCGGAGGTGAAGCTTCTATCAGGGCCA---GCG 497
Qy 683 GCTTGACGGCGCAGAAAGGCTCCGGCGCGGACGCTGCAACCGCTCGATCTCTCCCATCG 742
Db 498 GATGCCCGGAACGAGCGCGCTGAGCGCGCGCTGCAACCGCTGCTGTGGAGGCCAA 557
Qy 743 CGATGACCTGGCGCTGCTCAACGGGACCTCCGCCATGACCGGGATCGCGTGTGATGC 802
Db 558 GGAGGGCTTCGCGCTGTGTGAACGGGACACAGGCCATGTGCGCGGTGGGACCTGTCTCCA 617
Qy 803 TCACGCTTCGCCCATCTCGGCAACTGGCGGTGGGTTGACGGCGCTGTTCGCGAAATG 862
Db 618 GCTTCGGCGGAGTCCCTGCGCGACATCGCCGACGTCGCGGGGCCATGACGCTGAGGG 677
Qy 863 TCTGAGAGCCGGAACGAGGATGGGCGCGGACCTGTCCGACCTCTCGCGCGGATCCCGG 922
Db 678 GCTGTGGGAAGCCACAAGCCCTTCAATCTGAGATTACGACGCTCGCGCGCACCGGG 737
Qy 923 ACAGAGGACCGCGCAGCGAGCTCGCGCCCGTGGACGCGCAGCGCGGTGTGTCG 982
Db 738 CCAGAGGACGTCGCGCGCACCTCGCGCGCATCTGTGTGACAGGAGTGTGTGAGTC 797
Qy 983 GCACGTCAATTGCGAGCGGAGGCTCGACGCGCGCATATCGGACGAGCGGAGCGGG 1042
Db 798 GCACGTCA-----ACTGCAGCAAGT 818
Qy 1043 GCAGATGCTACAGCTCGCTCGCTCGCAGGTTCTGGGGCGGGCTTCGACACGCT 1102
Db 819 GCAGGACCTTCTCCCTCGCTGATGCGCGAGGTGACGCGCGCGCGCGAGGGCAT 878
Qy 1103 CGCATGGCATGACCGGCTGCTGACGATCGAGCTGAACCGGTGACGACAAATCCGGTGT 1162
Db 879 CGGTTCTCCCGCGCATCTTGAGGTGAGGTCAACAGCGCGACGACAAACCGCTCGT 938
Qy 1163 TCGCGCGATGGCAGCGTCCCGCTCGACCGGGGCAATTCATGGGCGAGCATGTGGC 1222
Db 939 ---GTTTCGGGACACGAGCGCATCGTGTGGGCGGCAACTTCCACGGCGAGCCATCTC 995
Qy 1223 GCTGAGCTCCGATGCGCTCGCCACGCGCTCACCGTTCTGGGGCGGCTTTCGAGAGCGCCA 1282
Db 996 CTTGGCATGAGCGTGTGGCGATGGCGCTGACGCAACTGTCTCATCAGCGAGCGGG 1055
Qy 1283 GATTGCACGCTGACAGATGAAGGCTGAACCGTGGCTGCCCGCTTCTCTCACCGGG 1342
Db 1056 CGTGGAGAGCTCGTGAAACCGCTCGCTGTCAAACCTGCGCGGT-----TCTCGGGA 1109
Qy 1343 CCGCGCGGCTTGAATTCGGGCTTCATGGGCGCACAGGTGACGCGGACCGCTCTCTGGC 1402
Db 1110 GAATCCGGGTTGAATCTCCGGCTTCATGATCGCGCAGGTGACGAGCGCGCTGTGGTGC 1169
Qy 1403 CGAGATGCGAGCCACGGGA---CCTGCGCTCGATTCGATTCGATTCACGAAACCGCGCCA 1459

Db 1170 CGAGTCCCGCGTGTGAGCCACCCGCGTCCGTGATTCGATTCGATCCCGGCGC 1229
Qy 1460 TCAGATGTGTGCTGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAAATGACGCG 1519
Db 1230 AGAGACCACTGTCCATGGGCATGACGCGCGCTCAAGGGCGCTCAGGTTCAGCACTT 1289
Qy 1520 TTGGGCGGAGATCTTGGATCTCGCTCTCTGCTCTGCTTTCGACAACTGCGGAGCTGCGCTG 1579
Db 1290 CGCCGCTTGTGCGCTTCGCGATGGAATCTTGGTGGCGCGCAGGCCCTTGGACTTCGCGCT 1349
Qy 1580 CGGACGCGGCTAGACCGGGTGTCTCCCGCGGGGAAAGCTGCTGAGGCGCTTGCAGG 1639
Db 1350 GCGCTGAAGCCCGCAAGGCGCTCTCGCGCTACGAGCTGG-----TGGCGTC 1400
Qy 1640 GCAATTCGCGCGCTTGAAGACGACCGGCGCTCTGGAAGAGAAATGCGCGCTTGTGTAC 1699
Db 1401 GAAGTCCCCACATGACAGGACCGAGCTGCACCGGACATCGAGGCGGTGAGCCA 1460
Qy 1700 GCACCTC 1706
Db 1461 GCTCGTC 1467

RESULT 13
US-10-369-493-28443
; Sequence 28443, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven S.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 28443
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-369-493-28443

Query Match 11.4%; Score 211.6; DB 6; Length 1503;
Best Local Similarity 51.3%; Pred. No. 1.4e-45;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

Qy 301 TCGGCGCTGTATCCGCGAGCGCGCATGTCTACGACTGACAAACCGGCTTCGGTCCCG 360
Db 125 TCGCGGATATGTCGCGGAGGCGGAGCGGCTACGCGATCAACACGCGCTTTGGGCGC 184
Qy 361 TTGCGAACCGCTGTATCTCAGGTGAGAAATGTCCGACGCTGACGGCAATCTTGTCCATC 420
Db 185 TCGCCAGCAGCATATTCGCGACGACCACTCGAACTGTTGACGCGCAATCTGTGTCT 244
Qy 421 ATCTGGCCAGCGGCTGGGACCGGTGCTGTGACTGGAACGCGCGCGCCCATGTTCTGG 480
Db 245 CTCACGCGGTGCGGTGGTGAGCCGATGTCCGCTCCGCGTGTGCTGTGATCGCGC 304
Qy 481 CGGCTGTGTGCTGATCGCTCAGGAGCGCTCCGCTGCCAGGAGGAGCCATCGCTGCC 540
Db 305 TGAACCTCTGAGCTTCGCGCGCGCCCAATTCGGGCAATTCGCGGCTGAAATGATGACGCGC 364
Qy 541 TGATCGACCTGCTCAATTCGAGCTCGCTCGGCGCTTCCAGCGCGGCGACGCTGGGCG 600
Db 365 TGATCAGCTGTACACGCGGACGCTGTGCTGCGGGTGAATTCGGTCAAGGTTCTGCTCGGTG 424
Qy 601 CGTCCGGTGACCTGACACCGCTTCGCGCATATGTGTCTCTGCTCCAGGCGCGGAGACT 660

RESULT 14
US-10-369-493-31202
; Sequence 31202, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei

Db 425 CATCGGCGACCTCGCGCGCTCGCGCATATGTCCGGCGCGCTCTCGGCGTCCGCGA-- 482
Qy 661 TCTTGGACCGGACCGGACCGGCTTTCAGCGCGAGAAAGGCTCCGGCGCGGACGCGTGC 720
Db 483 -AGTGTTCGGAAGGCGGAGCGCATGCGCGCCACCGAGGGTCTCGCGCTCTCGGCGCTCA 541
Qy 721 AACCGCTCGATCTCTCCATCGGATGCACTGGCGCTGCTCAACGCGGACCTCCGCCATGA 780
Db 542 AGCGCTCA CGCTGCAAGCCAAAGAGGGGCTGGCGCTCTCTGAAACGCGCACGCGAGCTTCGA 601
Qy 781 CCGGGATCGCGCTCGTGAATGCTCACGCCCTGCCCATCTCGGCAACTTGGGCGGTGGCGT 840
Db 602 CGCGCTCGCGCTGTACAAATGTTTCCCATCGAAGACCTGTATCCGACCGCGCTGTGT 661
Qy 841 TGACGGCGCTCTGTTGCGGAATGTCTGAGAGCCGGAACGAGGCAATGGGCGCGGACTGT 900
Db 662 CGGCGCATTTGTCGCTAGATCGCGCAATGGGCTCGGTCAAGCCGCTTCGACGCGCGCATTC 721
Qy 901 CCGACCTGCGGCGCGCATCCCGGACAGAGACCGCGCAGGAGGCTGGGCGCGCGCTGG 960
Db 722 ACAGTTTGC CGCGCATCAAGGTCAGATCGACGCGCGCGGCAATACCGCTCGCTGTGT 781
Qy 961 ACGCGACGCGCGGTGCTCGCGCACGTCATTTGCCGAGCGGAGGCTCGACGCGCGCGATA 1020
Db 782 AAGTTCCGCAATCAACGTTTCGCAAG----- 808
Qy 1021 TCGGGACGAGCGCGGAGCGCGGAGGATGCTTACAGCTTGCCTGCGCTCCCGAGTTC 1080
Db 809 -----CCGATTGCGACAAAGTGCAGGACCGGTACAGCTTGGCTGCCAGCGCAAGTCA 862
Qy 1081 TCGGGCGGCGCTTCGACACGCTGCATGCGATGACCGGGTGTCTGACGATCGAGCTGAACG 1140
Db 863 TGGGCGCGCTGTCTGGATCAGATCGCCACGCGGCAACGCTGCTGCTCGAAACGCAACG 922
Qy 1141 CGGTGACCGACAATCCGGTGTTCGCGCGCATGCGACGCTGCCCGCTTCACCGGGGCA 1200
Db 923 CGGTCTCCGACAATCCGCTGATTTCCCGA---CACCGGGAAGTGTGTTCGGGCGGTA 979
Qy 1201 ATTTTCATGGGCGCAGCATGTGCGCTGACGCTCCGATGCGCTTCGCCACGCGCGCTCAACGCTTC 1260
Db 980 ACTTTCACGACGAGCGGTTCGCTTCGCGCGCGATAACCTTCGCGCTCGCGCGCGCGAAA 1039
Qy 1261 TGGGCGGCTTTCGGAGCGCGCATGTCACGCTTCGACAGATGAAAGCTGNAACGCTGGGC 1320
Db 1040 TCGGCGCGCTTGGCGCAACGCGCATCGCTGCTGTGATCGACGCGACGCTGTTCGGGCGCTGC 1099
Qy 1321 TGCCCGCTTCTCCACCGCGGCGCGCGCTTGAATTCGCGCTTCATGGGCGCACAGG 1380
Db 1100 CGCC-----GTTCTCTGTCGCGATGCGCGCGTGAATTCGGGCTTCATGATCGCGCACG 1153
Qy 1381 TGACGCGCACCGCGCT---CCTGGCCGAGATGCGAGCCACGGGACCTGCTTCGATTCGATTC 1437
Db 1154 TCAGGCTCGCGCGCTCGCTTCGGAACAAAGACGCTCGCGCATTCGCGCTTCGCTCGATTC 1213
Qy 1438 CGATCTCCAGAACGCGCGCAATCAGGATGTGTCTGCTTGGGACCATCGCGCGGCGCC 1497
Db 1214 CGCTGCGCACTTCGGCAAAACGAGAAACACGCTGTGATGGGAGACGTTTCGCGCGGCGCA 1273
Qy 1498 TCTTCCCGGAGAGATCGACCGTTGGGCGGAGATCTTTCGATCTCTGCTCTCTGCTT 1557
Db 1274 AGTTCGCGCATATTCGTAACAAACCGCGCGCAATTCGTGATCGAACTGCTCGCGCGC 1333
Qy 1558 CACAAGCTCGGAGCTGCGC 1577
Db 1334 CGCAAGGCTCGATTCGCGC 1353

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31202
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31202

Query Match 11.4%; Score 211.6; DB 6; Length 1503;
Best Local Similarity 51.3%; Pred. No. 1.4e-45;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY 301 TCGGCGCTGTCATCCGAGGCGCGCATGTCTACGACTGACAAACGGCTTCGGTCCCC 360
Db 125 TCGCCGATATCGCCGGAAGGCGAGCCGCCCTACGGCATCAACACGGGCTTTGGGCGCC 184
QY 361 TTGCGAACCCCTGTATCTCAGGTGAGAAATGTCCGAACGCTGCAGGGCAATCTTGTCCATC 420
Db 185 TCGCCAGCAGCATATTCGCGACGACCACTCGAATCTGTGAGCGCATCTGGTGTCT 244
QY 421 ATCTGCGCAGCGCGTGGGACCGGTGTTGACTGGAACGCGCGCGCCATGGTTCG 480
Db 245 CTCAACGCGTGGGCTGGGTGAGCCGATGTCGGTCCGGTGGTCTGCTGATCGCG 304
QY 481 CGCGTGTGTTGATCGCTCAGGAGGCTCCGGTGCAGGAGGGAACATCGCTCGCC 540
Db 305 TGAACCTCTCGAGCCTCGCGCGCGCCATTCGGGCAATTCGGCTGAAGTATGACGCGC 364
QY 541 TGATCGACCTGTCAATTCGAGCTCGCTCGGCGCTTCCAGCGCGGACGGTGGCG 600
Db 365 TGNATCGCTGTACACGCGAGCTGCTCGCGTGAATTCGGTCAAGGTTCCGTCGGTG 424
QY 601 CGTGGGTGACTGACACCGCTTTCGCAATGATGCTCTGCTCCCTCCAGGCGCGGGAGACT 660
Db 425 CATCGGCGACCTCGCGCGCTCGCATATGTCGGCGCGCTGCTCGCGTGGGGA-- 482
QY 661 TCTTGACCGGACGGAACGCGCTTGAACGCGCAGAGGCTCCGGCGGAGCGCTGC 720
Db 483 -AGTGTTCGGAAGGCGAGCGATCCGCGCCACCGAGGTTCTCGCGCTCGTGGCTCA 541
QY 721 AACCGCTCGATCTCTCCATCGGATGCACTGGCGCTGTGCAACGGGACCTCGCCATCA 780
Db 542 AGCGCTCAGCTGCAAGCAAGAGGGGTGGCTCTCTGAACGGCAGCGAGCTTCA 601
QY 781 CCGGGATCGCGCTGTGTAATGTCTACGCTTCGCGCATCTCGGCAACTGGGCGGTGGCT 840
Db 602 CGCGCTCGCGCTGTACAACTGTCGCCATCGAAGACTGTACCCGACCGCGCTGGT 661
QY 841 TGACGCGCTCTGTTGGGAATGTCTGAGAGCGCGGACCGAGGATGGGCGCGGCACTGT 900
Db 662 CGGGCGCATTTGTCGTTAGATGCGCAATGGGCTCGGTCAAGCGCTTCGACGCGCGCATTC 721
QY 901 CGACCTCGCGCGCATCCGAGACAGAGCGCGCAGCGAGGCTGCGCGCGCGCTGG 960
Db 722 ACGAGTTGCGCGCATCAAGTTCAGATCGACCGCGCGGCGCATACCGCTCGTCTGG 781
QY 961 ACGGACGCGCGGTTGGTTCGCGCACTGTTCCGAGCGAGGCTTCGACGCGCGGATA 1020
Db 782 AAGTTTCGCAATCAACGTTTCGACG----- 808
QY 1021 TCGGACGAGCGGAGGCGGGCAGGATGCTACAGCTGCTGCTGCGCTCCGAGGTTTC 1080
Db 809 -----CCGATTTCGCAAGAGGTGACAGGACCCGTTACAGCTTGGCTTCGACGCGCAAGTCA 862

RESULT 15

US-10-282-122A-13653
; Sequence 13653, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13653
; LENGTH: 1521
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13653

Query Match 11.4%; Score 211.6; DB 7; Length 1521;
Best Local Similarity 51.3%; Pred. No. 1.4e-45;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY	301	TCGGCGCTGTCATCGCGGAGCGGGCCATGCTACGGACTGACACCGGGTTCCGTCGCC	360
DB	122	TCGCCGATATCGCGCGAAGGGGAGCGGCCCTACGGCATCAACACGGGGCTTTGGGGGCC	181
QY	361	TTCCGAACCGCTGATCTCAGGTGAGATCTCCGAACGCTGCAGGCCCAATCTTGTCCATC	420
DB	182	TCGCCAGCAGCATATTCGCCACGACCAACTCGAACTGTTGACGGCAATCTGGTCTCT	241
QY	421	ATCTGGCCAGCGCGTGGGACCGGTGCTGACTGGACGACGCGCGCCGTCATGGTTCTGG	480
DB	242	CTCACCGGTCGGCGTGGGTGAGCCGATGTCGGTCCGGTCGGTCTGCTGATCGCGC	301
QY	481	CGCGTCTGGTGTGATGCTCAGGAGACCTTCGGTGCAGGAGGACCATCGCTCGCC	540
DB	302	TGAACCTCTCGAGCCTCGGCGCGGCCCATTCGGGCAATTCGCGTGAAGTATGAGCGCGC	361
QY	541	TGATCGACCTGCTCAATTCGAGCTCGCTCCGCGCGTTCCACGCGCGGCACTGGGG	600
DB	362	TGATCAGCTGTACACGCGCGAGTGTGCTCGGTGATTCGGTCAAGGGTTCCGTCGGTG	421
QY	601	CGTCCGGTGACCTGACACCGCTTTCGCATATGGTGTCTTCCCTCCAGGGCCGGGAGACT	660
DB	422	CATCGGGCGACCTCGCGCGCTCGCATATGTCGGCGCGCTCTCGGGTCGGCGA--	479
QY	661	TCCTGACCGGGAGGGAGCGCGCTTGAACGGCGAGAGGGCTCCGGCGCGGACGGCTGC	720
DB	480	-AGTGTTCGGAAGGGGAGCGATGCGGCCACCGAGGGTCTCGCGCTCGTCCGGCTCA	538
QY	721	AACCGCTCGATCTCTCCATCGCGATGCACTGGGCTGGTCAACGGGACCTCCGCCATGA	780
DB	539	AGCCGCTCAGCTGCAAGCCAAAGAGGGGCTGGCGCTCTCTGAACCGGACGCGAGCTTCGA	598
QY	781	CCGGGATCGCGCTGGTGAATGCTCAGCCCTCGCGCCATCTCGGCAACTGGGCGGTGGCGT	840
DB	599	CGCGCTCGCGCTGTACACATGTTGCCCATCGAAGACCTGTACCGCAACGCGCTGGTGT	658
QY	841	TGACGCGCTCTGTTGCGGAATGCTGAGAGCGCGAACCGAGGCAATGGGCGCGGCACTGT	900
DB	659	CGGGCGCATTTGTCGGTAGATGCGGCAATGGGCTCGGTCAGCGGCTTCGACGCGCGCATC	718
QY	901	CCGACCTGGCGCGCATCCCGGACAGAGACCGCGAGGCTTCGGCGCGCGGTGG	960
DB	719	ACGAGTTGGCGGCCCATCAAGGTCAGATCGACGCGCGCGGCAATCCGCTCGCTGTGG	778
QY	961	ACGCGAGCGCGCGGTGGTTCGGGCAAGTCAATTCGCCGAGCGAGGGCTCGAGCGCGCGGATA	1020
DB	779	AAGTTTCGGCAATCAACGTTTTCGACG-----	805
QY	1021	TCGGGACGAGCGGAGCGGGGAGAGATGCCCTACAGCTTGGCTGCGCTCCGACGGTTC	1080
DB	806	-----CCGATTGCGACAGGTCGAGGACCGGTACAGCTTGGCTGCGCTCGAGCCGCAAGTCA	859
QY	1081	TCGGGGGGGGCTTCGACACGCTCGCATGGCATGACCGGGTGTGTCAGCATCGAGCTGAACG	1140

DB	860	TGGGCGGTGCTCTGGATCAGATCGGCACGCGGCCAACGTGCTGCTCGAAGCGAACG	919
QY	1141	CGGTGACCGACAATCCGGTGTTCGCCGCCGATGGCAGCGTGGCCGCTTGACACGGGGCA	1200
DB	920	CGGTCTCCGACAATCCGTGATTTTCCCGA---CACCGCGAAGTGTCTGTCGGGCGGTA	976
QY	1201	ATTTTCATGGGCCAGCATGTGGCGCTGACGTCGATGGCTCGCACGCGGCTCACCGTTC	1260
DB	977	ACTTCCACGAGAGCGCGTCCCTTCGCGGCCGATTAACCTCGCGCTCGCGCGCCGAAA	1036
QY	1261	TGGCGGCGCTTTCGGAGCGCAGATTGACGCTCTGACAGATGAAAGCTGAACCGTGGGC	1320
DB	1037	TCGGCGCGCTGGCGCAACCGCGCATCGCGTGTGATCGACGCGACGCTGTTCGGGCTGC	1096
QY	1321	TGCCCCCTTCTCCACCGGGGCCCGCGGTTGAATTCGGCTTCATGGGCGCACAGG	1380
DB	1097	CGCC-----GTTCTCTCGTGGCGATGGCGCGTGAATTCGGGCTTCATGATCGCGACG	1150
QY	1381	TGACGGGAGCGCGCT---CCTGGCCGAGATCGAGCCACGGGACCTGCTTCGATCCATT	1437
DB	1151	TCACGGCTCGGGCGCTCGCCTCGGAAACAAAGACGCTCGCGCATTCGGCTTCGGTTCGATT	1210
QY	1438	CGATCTCCAGCAACGCGCCCAATCAGGATGTGTCTCGCTTGGACCATCGCCGCGGCC	1497
DB	1211	CGTTCGCCACTTCGGCAACCCAGGAAGACCACTGTGATGGCGACGTTGCGCGCGCA	1270
QY	1498	TTTCGGCGGAGAAAGATCGACCGTTGGGCGGAGATCCTTTCGATCTCTGCTCTTGT	1557
DB	1271	AGCTCGGCGACATTTGCTGAAACACACCGCAACATTTCTGTCGATCGAACTGCTCGCCGCGC	1330
QY	1558	CACAAGCTCGGAGCTGGGC 1577	
DB	1331	CGCAAGCGCTCGATCTCGCGC 1350	

Search completed: December 10, 2005, 12:12:48
Job time: 1573.secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2005, 08:52:45 ; Search time 1129 Seconds

(without alignments)
10920.887 Million cell updates/sec

Title: US-10-621-826-2

Perfect score: 1850

Sequence: 1 ccgagggggccatccacgag.....tcagctcgatcccgacgagg 1850

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_21.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*
14: geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1850	100.0	1850	12	ADJ57135		Adj57135 R. sphaer
2	1446	78.2	1446	13	ADS55829		AdS55829 Bacterial
3	561	30.3	1626	12	ADL16876		Adl16876 Rhodobact
4	244.2	13.2	1587	13	ADS56825		AdS56825 Bacterial
5	242.4	13.1	42000	3	AA63349		Aa63349 Streptomy
6	242.4	13.1	63164	3	AA63348		Aa63348 Streptomy
7	238.4	12.9	1497	13	ADS55996		AdS55996 Bacterial
8	231.6	12.5	1521	8	ACA25972		ACA25972 Prokaryot
9	215.4	11.6	1503	13	ADT44460		Adt44460 Bacterial
10	215.4	11.6	1527	14	ACL71540		ACL71540 M. xanthu
11	215.4	11.4	6439	14	ACL64350		ACL64350 Bacterial
12	211.6	11.4	1503	13	ADS50013		AdS50013 Bacterial
13	211.6	11.4	1521	8	ACA25783		ACA25783 Prokaryot
14	211.6	11.4	1521	8	ACA25783		ACA25783 Prokaryot
15	177.4	9.6	1518	13	ADS50000		AdS50000 Bacterial
16	177.4	9.6	1518	13	ADS55516		AdS55516 Bacterial
17	177.2	9.6	1575	13	ADT48701		Adt48701 Bacterial
18	173.4	9.4	1530	4	AA54350		Aa54350 Pseudomon
19	173.4	9.4	1530	8	ACA42730		ACA42730 Prokaryot

20	173.4	9.4	1530	13	ADT42089		Adt42089 Bacterial
21	172.8	9.3	34980	6	ABQ81847		Abq81847 Bifidobac
22	172.8	9.3	34980	6	ABQ81848		Abq81848 Bifidobac
23	171.8	9.3	1548	11	ABD17871		Abd17871 Pseudomon
24	171.8	9.3	2295	11	ABD17522		Abd17522 Pseudomon
25	170.8	9.2	1530	13	ADS59513		AdS59513 Bacterial
26	170.8	9.2	1533	13	ADS62488		AdS62488 Bacterial
27	170.8	9.2	1533	13	ADS63119		AdS63119 Bacterial
28	170.8	9.2	1542	13	ADS60352		AdS60352 Bacterial
29	163.6	8.8	30390	4	AA559520		Aa559520 Propionib
30	163.2	8.8	30390	8	ACF64449		Acf64449 Propionib
31	163.2	8.8	1527	11	ABD00629		Abd00629 Klebsiell
32	161.8	8.7	1542	6	AAD26199		Aad26199 Corynebac
33	161.6	8.7	1521	13	ADS63795		AdS63795 Bacterial
34	161.6	8.7	1521	13	ADS64178		AdS64178 Bacterial
35	161.6	8.7	1521	13	ADS63427		AdS63427 Bacterial
36	159.4	8.6	1566	13	ADS57133		AdS57133 Bacterial
37	157	8.5	1525	8	ACA35512		ACA35512 Prokaryot
38	155.2	8.4	1545	8	ACA45293		ACA45293 Prokaryot
39	153.8	8.3	1536	6	AA262200		AA262200 Corynebac
40	151.8	8.2	1446	13	ADS50081		AdS50081 Bacterial
41	151.8	8.2	1491	13	ADS55596		AdS55596 Bacterial
42	151.6	8.2	1452	13	ADS56654		AdS56654 Bacterial
43	149.4	8.1	1530	13	ADS61703		AdS61703 Bacterial
44	146.6	7.9	1530	8	ACA44103		ACA44103 Prokaryot
45	144.6	7.8	1461	13	ADS49545		AdS49545 Bacterial

ALIGNMENTS

RESULT 1

ADJ57135
ID ADJ57135 standard; DNA; 1850 BP.

XX AC ADJ57135;

XX DT 06-MAY-2004 (first entry)

XX DE R. sphaeroides tyrosine ammonia lyase (TAL) enzyme encoding DNA.

XX KW TAL; tyrosine ammonia lyase; catalyst; para-hydroxycinnamic acid; PHCA;

XX KW tyrosine; gene; ds.

XX OS Rhodobacter sphaeroides.

XX FH Key Location/Qualifiers

XX FT CDS 147..1727

XX FT /*tag= a

XX FT /product= "TAL"

XX FT /note= "tyrosine ammonia lyase"

XX FT /partial

XX FT /note= "the start codon is not indicated"

XX FN WO2004009795-A2.

XX PD 29-JAN-2004.

XX PF 23-JUL-2003; 2003WO-US023229.

XX PR 23-JUL-2002; 2002US-0397820P.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Huang L, Xue Z;

XX PFI WPI; 2004-214331/20.

XX P-ESDB; ADJ57136.

XX PT Novel tyrosine ammonia lyase enzyme, used as catalyst in producing para-

XX PT hydroxycinnamic acid from tyrosine.

XX PS Claim 2; SEQ ID NO 2; 53pp; English.

xx The invention relates to a tyrosine ammonia lyase enzyme (TAL) and
cc encoding polynucleotides. The TAL polypeptide is used as a catalyst in
cc producing para-hydroxycinnamic acid (PHCA) from tyrosine. The method
cc involves contacting a transformed host cell expressing the polypeptide
cc with a fermentable carbon substrate, e.g. monosaccharides,
cc oligosaccharides, polysaccharides, carbon dioxide, methanol,
cc formaldehyde, formate and carbon-containing amines, preferably glucose,
cc growing the host cell for a time sufficient to produce PHCA, and
cc optionally recovering PHCA. The TAL polypeptide enables simple and cost-
cc effective production of PHCA. The present sequence represents a R.
cc sphaeroides TAL polypeptide encoding DNA.
xx
SQ Sequence 1850 BP; 282 A; 628 C; 635 G; 305 T; 0 U; 0 Other;

Query Match 100.0%; Score 1850; DB 12; Length 1850;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGAGGCGCCATCCACGAGATGGCGCCGCTTCGGGCTGACGCCCGCGGATCTCTGC 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 61 TCGATCTGGGCGCGCTGGGCGGACGATCTGAAGCAAGAACTCTGCATTAAACAGCTTC 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 61 TCGATCTGGGCGCGCTGGGCGGACGATCTGAAGCAAGAACTCTGCATTAAACAGCTTC 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 CACTACCTTCGGCGAGAGACAGGAGTGAAGCCAAATGCTCGCCATGAGCCCGCGAAGC 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 121 CACTACCTTCGGCGAGAGACAGGAGTGAAGCCAAATGCTCGCCATGAGCCCGCGAAGC 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181 CGGCGCTGAGCTGATCGCACATCGATCTGGACCGGCCATGCGTGGCGGAGCGCG 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 181 CGGCGCTGAGCTGATCGCACATCGATCTGGACCGGCCATGCGTGGCGGAGCGCG 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 241 CGCGCGGATTTGTCCTTCCCTTCGGCGCGGACCGGTGCGGTCCGAGCGCGG 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 241 CGCGCGGATTTGTCCTTCCCTTCGGCGCGGACCGGTGCGGTCCGAGCGCGG 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 301 TCGGGCTGTGTCATCGCGAGGCGGCCATGCTACGAGCTGACAAACCGGCTCGTCCCC 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 301 TCGGGCTGTGTCATCGCGAGGCGGCCATGCTACGAGCTGACAAACCGGCTCGTCCCC 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 361 TTGGAAACCGCTGATCTCAGTGAGATGTCGAAAGCTGACGCGCAATCTGTTCATC 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 361 TTGGAAACCGCTGATCTCAGTGAGATGTCGAAAGCTGACGCGCAATCTGTTCATC 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 421 ATCTGGCAGCGGCTGGGACCGGTGCTGATCGAACGCGCGCGGCCATGTTCTGG 480
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 421 ATCTGGCAGCGGCTGGGACCGGTGCTGATCGAACGCGCGCGGCCATGTTCTGG 480
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 481 CGGCTGTGTTGATCGCTCAGGAGGCTCGGTGCGGAGGAGGACCATCGCTCGCC 540
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 481 CGGCTGTGTTGATCGCTCAGGAGGCTCGGTGCGGAGGAGGACCATCGCTCGCC 540
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 541 TGATCGACTGCTCAATTCGAGCTCGCTCGGCGCTTCCAGCGCGGCGACGCTGGGCG 600
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 541 TGATCGACTGCTCAATTCGAGCTCGCTCGGCGCTTCCAGCGCGGCGACGCTGGGCG 600
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 601 CGTGGGCTGACCTGACACCGCTTGGCATATGTTGCTCTGCTCCAGGCGCGGAGACT 660
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 661 TCCTGGACCGGAGACGGGCTTGACGGCGCAGAAAGGCTCGGCGGAGCGGCTG 720
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 661 TCCTGGACCGGAGACGGGCTTGACGGCGCAGAAAGGCTCGGCGGAGCGGCTG 720
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 721 AACCGCTGATCTCTCCATCGCATGCACTGGGCTGTGTCATCGGAGCTTCGCCATGA 780
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 721 AACCGCTGATCTCTCCATCGCATGCACTGGGCTGTGTCATCGGAGCTTCGCCATGA 780
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 781 CCGGGATCGCGCTGGTGAATGCTCACGCTCGCCCATCTCGGCAACTCGGCGGTGGCGT 840
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

AD855829 standard; cDNA; 1446 BP.
AD855829;
02-DEC-2004 (first entry)
Bacterial polynucleotide #7816.
Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.
Bacteria.
US2003233675-A1.
18-DEC-2003.
20-FEB-2003; 2003US-00369493.
21-FEB-2002; 2002US-0360039P.
(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.
Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS; WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.
Claim 1; SEQ ID NO 31503; 122pp; English.
The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.
Sequence 1446 BP; 212 A; 487 C; 505 G; 242 T; 0 U; 0 Other;
Query Match 78.2%; Score 1446; DB 13; Length 1446;
Best Local Similarity 100.0%; Pred. No. 3.2e-259;
Matches 1446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 GACCGTGCCTGCTGCCAAGCGCGCTCGCGCTGTCTATCCGAGGCGCGCATGTC 332
DB 1 GACCGTGCCTGCTGCCAAGCGCGCTCGCGCTGTCTATCCGAGGCGCGCATGTC 60
QY 333 TACGGA CTGACAACCGGCTTTCGCTTCCCTTTCGAAACCGCTGTCTCAGGTGAGAAATGC 392
DB 61 TACGGA CTGACAACCGGCTTTCGCTTCCCTTTCGAAACCGCTGTCTCAGGTGAGAAATGC 120
QY 393 CGAAGCTGCAGGCGCAATCTTGTCTCATCTATGCGGCGGCTGGGACCGGTGCTTGAC 452
DB 121 CGAAGCTGCAGGCGCAATCTTGTCTCATCTATGCGGCGGCTGGGACCGGTGCTTGAC 180
QY 453 TGGACGACGGCGCGCGCATGTTCTGCGCGCTCTGTGTGATCGCTCAGGAGCCCTCC 512
DB 181 TGGACGACGGCGCGCGCATGTTCTGCGCGCTCTGTGTGATCGCTCAGGAGCCCTCC 240
QY 513 GGTGCCAGCGAGGGGACCATCGCTCGCTGTATCGACCTGTCTCAATCCGAGCTCGCTCCG 572
DB 241 GGTGCCAGCGAGGGGACCATCGCTCGCTGTATCGACCTGTCTCAATCCGAGCTCGCTCCG 300
QY 573 GCGTTTCCAGCGCGGCGCATGCTGGGCGGCTCGGCTGACCTGACACCGCTTGGCATATG 632
DB 301 GCGTTTCCAGCGCGGCGCATGCTGGGCGGCTCGGCTGACCTGACACCGCTTGGCATATG 360
QY 633 GTGCTCTGCTCCAGGGCGCGGAGACTTCTTGACCGGGACGCGCGCTTGGACGGC 692
DB 361 GTGCTCTGCTCCAGGGCGCGGAGACTTCTTGACCGGGACGCGCGCTTGGACGGC 420
QY 693 GCAGAAGGGCTCCCGCGCGGACGCGCTGCAACCGCTCGATCTCTCCCATCGGATGCACTG 752
DB 421 GCAGAAGGGCTCCCGCGCGGACGCGCTGCAACCGCTCGATCTCTCCCATCGGATGCACTG 480
QY 753 GCGTGTGTAACGGGACCTCGCCCATGACCGGGATCGGCTGTGTGTAATGCTCAGCCCTGC 812
DB 481 GCGTGTGTAACGGGACCTCGCCCATGACCGGGATCGGCTGTGTGTAATGCTCAGCCCTGC 540
QY 813 CGCCATCTCGGCACTCGGCGGTGCGCTGACCGCCCTGCTTTCGGAATGCTCAGAGGC 872
DB 541 CGCCATCTCGGCACTCGGCGGTGCGCTGACCGCCCTGCTTTCGGAATGCTCAGAGGC 600
QY 873 CGGACCGAGGCTCGGCGCGGCACTGTCGACCTCGCGCGCATCCCGGACAGAAGGC 932
DB 601 CGGACCGAGGCTCGGCGCGGCACTGTCGACCTCGCGCGCATCCCGGACAGAAGGC 660
QY 933 GCGCAGCGAGGCTCGCGCGCGGCTGGAACGGCAGCGCGGCTGCTTCGGCAGCTCAT 992
DB 661 GCGCAGCGAGGCTCGCGCGCGGCTGGAACGGCAGCGCGGCTGCTTCGGCAGCTCAT 720
QY 993 GCGCAGCGAGGCTCGGACCGCGGATATCGGACCGGCGCGGAGCGGCGGCGAGGATGCC 1052
DB 721 GCGCAGCGAGGCTCGGACCGCGGATATCGGACCGGCGGAGCGGCGGCGAGGATGCC 780
QY 1053 TACAGCTCGGCTCGCTCGCTCGCGAGGCTTCGCGCGGCTTCGACACGCTCGCATGGCAT 1112
DB 781 TACAGCTCGGCTCGCTCGCTCGCGAGGCTTCGCGCGGCTTCGACACGCTCGCATGGCAT 840
QY 1113 GACCGGCTGTGACGATCGAGCTGAAACGGCTGAAACGCAATCCGGTGTTCGCGCGAT 1172
DB 841 GACCGGCTGTGACGATCGAGCTGAAACGGCTGAAACGCAATCCGGTGTTCGCGCGAT 900
QY 1173 GCGAGCTGCGCGCGCTCGACCGGGGCAATTCATGGCGGAGCATGTGGCGCTGACGCTCC 1232
DB 901 GCGAGCTGCGCGCGCTCGACCGGGGCAATTCATGGCGGAGCATGTGGCGCTGACGCTCC 960
QY 1233 GATCGCTCGCGACCGCGCTCAGCGTTCGCGGGCTTCGCGGAGCGCGAGATTCGACGT 1292
DB 961 GATCGCTCGCGACCGCGCTCAGCGTTCGCGGGCTTCGCGGAGCGCGAGATTCGACGT 1020
QY 1293 CTGACAGTGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGGGCCCCCGGG 1352
DB 1021 CTGACAGTGAAGGCTGAACCGTGGGCTGCCCCCTTCTCCACCGGGGCCCCCGGG 1080
QY 1353 TTGAATTCGGCTTCATGGGCGCACAGGTGACGGCGCACCGGCTCTCTGGCGGAGATGGA 1412

Db 1081 TTTGAATTCGGCTTCATGGCGCACAGGTACGGCGACCGCGCTCTCTGGCGGAGATCGGA 1140
Qy 1413 GCCACGGAGCTGCTCGATCCATTGCTCTCCAGCAAGCGCGCAATCAGGATGTGGTC 1472
Db 1141 GCCACGGAGCTGCTCGATCCATTGCTCTCCAGCAAGCGCGCAATCAGGATGTGGTC 1200
Qy 1473 TCGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCGTTGGCGGAGATC 1532
Db 1201 TCGCTTGGGACCATCGCGCGCGCTCTGCGCGAGAGATCGACCGTTGGCGGAGATC 1260
Qy 1533 CTTGCGATCTCGCTCTCTGTCTGTGCAAGCTCGGAGCTGCGCTCGCGAGCGGCTTA 1592
Db 1261 CTTGCGATCTCGCTCTCTGTCTGTGCAAGCTCGGAGCTGCGCTCGCGAGCGGCTTA 1320
Qy 1593 GACGGGTGTCTCCGCGGGGAGAGCTGTGTGAGGAGCTGTGAGGAGCTGTGCGGAGCTTCCCGCG 1652
Db 1321 GACGGGTGTCTCCGCGGGGAGAGCTGTGTGAGGAGCTGTGAGGAGCTGTGCGGAGCTTCCCGCG 1380
Qy 1653 CTTGAGACGACCGCGCCCTGGGACAGGAAATTCGCGCGCTTGTACGACCTCTTTGAG 1712
Db 1381 CTTGAGACGACCGCGCCCTGGGACAGGAAATTCGCGCGCTTGTACGACCTCTTTGAG 1440
Qy 1713 CAATCT 1718
Db 1441 CAATCT 1446

RESULT 3
ADL16876
ID ADL16876 standard; DNA; 1626 BP.
XX
AC ADL16876;
XX
DT 06-MAY-2004 (first entry)
XX
DE Rhodobacter capsulatus tyrosine ammonia lyase (tal) DNA.
XX
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; gene; ds.
XX
OS Rhodobacter capsulatus.
XX
FH Location/Qualifiers
FT 1..1626
FT /*tag=a
FT /product= "Rhodobacter capsulatus tyrosine ammonia lyase
FT (TAL) protein"
FT /transl_except= (pos:1608..1608, aa:Ser-Ala)
XX
PN US2004029230-A1.
XX
PD 12-FEB-2004.
XX
PF 18-JUN-2003; 2003US-00464609.
XX
PR 18-JUN-2002; 2002US-0389593P.
XX
PA (KIND/) KYNDT J J A.
PA (BEEU/) BEEUMEN J V.
XX
PI Kyndt JJA, Beeumen JV;
XX
DR WPI; 2004-168893/16.
DR P-PSDB; ADL16877.
XX
XX New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
XX synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
XX protein used as photochromic substance in electro optical random access
XX memory or in biochips.
PS Claim 21; SEQ ID NO 3; 32pp; English.
XX

CC The present invention is generally related to recombinant DNA technology
CC and particularly to DNA strands useful for the production of
CC parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
CC suitable host expression system. The invention is useful for synthesizing
CC parahydroxycinnamic acid and derivative products such as holo-photoactive
CC yellow protein which is useful as a photochromic substance in electro
CC optical random access memory, in biochips, in light-inducible gene
CC expression systems and in liquid crystal polymers. The present sequence
CC is Rhodobacter capsulatus tyrosine ammonia lyase (tal) DNA.
XX
SQ Sequence 1626 BP; 229 A; 528 C; 610 G; 256 T; 0 U; 3 Other;
Query Match 30.3%; Score 561; DB 12; Length 1626;
Best Local Similarity 60.9%; Pred. No. 9.9e-99;
Matches 947; Conservative 0; Mismatches 603; Indels 6; Gaps 2;
Qy 175 CGAAGCGGCGCTGAGCTGGATCGGCACATCGATCTGACACGAGCCCATCGCGTGGCGA 234
Db 53 CCAAGGACTGCTCTCGCGCTGGACGGGCGCTGACACTTGTCCAATCGAGGCCATCGCGA 112
Qy 235 GCGGCGCGCGCGGATTTGTCTTTCCTCCGCGCGCGACCGCTGCGTGCCTGCCAAG 294
Db 113 CACATCGACCGCGATTTGGTGACCCCGCGCTGCGGAGCGCTGCGCGGGCCCATG 172
Qy 295 GCGGCTCGGCGCTGTATCATCGCGAGCGCGCATGTCTACGAGCTGACAAACCGGTTTCG 354
Db 173 CCGGCTTGAGCACGCCATCGCGAGCAGCGCCACATTTACGGCATCACACCGGCTTCG 232
Qy 355 GTCCCTTGGAAACCGGCTGATCTCAGGTGAGAAATGTCCGAACGCTGACGCCAATTTTG 414
Db 233 GCCCGCTGGCGAAACCGTCTGATCGGGGCGGATCAGGGGCGGAGCTGACACGAAACCTGA 292
Qy 415 TCCATCATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGGAGCAGCGCGCGCGCCATGG 474
Db 293 TCTATCATCTGGCCAGCGCGTGGGCGGCGGAACTGAGCTGGGCGGAGCGCGCGGCTTGA 352
Qy 475 TTCTGGCGGCTCTGTGTGATCGCTCAGGAGCCTCCGCTGCGAGCGAGGGAACCATTCG 534
Db 353 TGTGCGGCGGCTCAACTGATCTCTCAAGCGCGCTCGGGGCGCTCGCCGAGACGATCG 412
Qy 535 CTGCGCTGATCGACCTGCTCAATTCGAGCTCGCTCGGCGGCTTCCAGCGCGGCGACGG 594
Db 413 ACCGATCTGTTGCGGTGCTCAATTCGCGGGTTCCTCCCGGAGGTTCCGCGCGACGGAACGG 472
Qy 595 TGGGCGGCTCGGCTGACCTGACACCGCTTCGCGCATATGGTGCTCTGCTCCAGGCGCGG 654
Db 473 TGGGCGGCTCGGCGGATCTGACCCCGCTTGGCGATATGTTGCTGCGGCTCGAGGACGG 532
Qy 655 GAGACTTCTTGGACCGGCGAGCGGCTTGAACGCGCGAGAGGCTCCGCGCGGAGC 714
Db 533 GCGGATGATCGACCCCTCGGGCGCGCTGCGAGGCGCGGGCGGTGATGGATCGGCTCT 592
Qy 715 GGCTGCAACCGCTCGATCTCTCCATCGGATGACATGGGCGCTGGTCAAGGGGACCTCCG 774
Db 593 GCGGCGGCTCGCTGACGCTGGCGCGCGCTGACGGGCTGGGCTGGTGAATGGCACCTCGG 652
Qy 775 CCATGACCGGGATCGCGCTGTTGATGCTCACGCCCTGCCGCCATCTCGGCAACTGGGCGG 834
Db 653 CGATGACCGGGATGCGGCGCTGACCGGGGTGAGGGGCGCGGGCGATCGACGCGCGC 712
Qy 835 TGGCGTTGACGCGGCTGCTTTCGGAATGTCTGAGAGCGCGGACCGAGGCTATGGGCGCGG 894
Db 713 TTGCGCACAGCGGGTCTCTGATGAGGCTTTGTCCGCTCATGCGGAAGCTTGGCATCCGG 772
Qy 895 CACTGTCCGACTGGCGGCGCATCCGAGCAGAGGAGCGCGCGGAGGCTGCGGCGCC 954
Db 773 CTTTTCGACAGCTGCGCGCGCATCCGCGGCGAGCTGCGGGCGGACCGAGCGGCTGCGCGAGG 832
Qy 955 GCGTTCGACGCGACGCGCGGCTGCTCCGCACTGCTATTTGCCGAGCGGAGGCTCGACCGCG 1014
Db 833 CGCTGATGGGCGGCGGGGCTCTGTGCGACCTTGACCGCGCGCGGCGGCTGACCGCGG 892
Qy 1015 GCGATATCGGGACGAGCGCGGCGGAGGATGCTTACGCTGCGCTGCGCTCCGCTCCG 1074

```
Db 893 CGGATCTGGGCCCCGAAGATCATCCGGCGCAGGATGCTTACAGTCTGGCGGTGTGCGCG 952
Qy 1075 AGTTTCTGGGGCGGGCTTCGACACGCTCGCATGATGACCGGGTGTGAGATCGAGC 1134
Db 953 AACTGGTGGGCGGCTCTGGGACACGCTGGAGTGGCAGATCGTGTCTGCTCACTCGGAGC 1012
Qy 1135 TGAACGGGTGACGACAAATCGGTGTTTCCGCCCGATGGCAGGTGCGCGCTTCGACG 1194
Db 1013 TCAATTCCGTCAACGACAAATCGATCTTTCCGAGGGCTGCGCGTGCCTCCCGCTGCACG 1072
Qy 1195 GGGGCAATTTATGCGGCGCAGATGCGGCTGACGTCGATGCGCTCGCCACGCGCGTCA 1254
Db 1073 GCGGCAATTTATGCGGCGTGAATGTCGCTTGCCTCGATGCGCTGAAACGCGCGCTGG 1132
Qy 1255 CCGTTCTGGGCGGCTTCGGGAGCGCCAGATGTCACGCTGTCACAGATGAAGGCTGAAC 1314
Db 1133 TGAAGCTGGGCGGCTGGTGGAGCGTCAATGCGCGGCTGACCGACGAAAAGCTGAACA 1192
Qy 1315 GTGGGCTGCGGCTTCTTCCACCGGGGCGCGCGGTTGAATTCGGGCTTCATGGCG 1374
Db 1193 AGGGCTGCGGCTTCTTCCGCGGCGGCGAGCGGGGCTGCAATCGGGCTTCATGGGG 1252
Qy 1375 CACAGGTGACGCGGACCGGCTCTTGGCGGAGATGCGAGCCA---CGGGACCTGCTCGA 1431
Db 1253 GCGAGGTCAAGGCGGACCGGCTTCTGGCGGAAATGCGGGCGAATGCCACGCGGTTTCGG 1312
Qy 1432 TCCATTTCGATCTCCACGAAACCGCGCAATCAGGATGTGCTCTCGTTGGGACCATGCGCG 1491
Db 1313 TGCAGTCTGCTCGACCAATGGCGCAATCAGGATGTGCTCGATGGAACGATTCGCG 1372
Qy 1492 CGCGCTCTGCGCGGAGAAATGCAACCTTGGGGGAGATCTTTCGATCTCGCTCTCT 1551
Db 1373 CGCGAGGGCGCGGCGGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1432
Qy 1552 GTCTTGACAGCTGCGGAGTGGCTGCGCGCAGC---GGCTAGACGGGTGCTCCG 1608
Db 1433 CCTTGGCCAGGATGATCTGTTGACAGCCCGGAGGGGAGCGCGGATGCTGCTGCTTA 1492
Qy 1609 CGGGGAAGAAGCTGTGTCAGGCTTTCGCGAGCAGATTCCTCCCGCTTGAGACGACCGCG 1668
Db 1493 CGCGCGGATCTGCGGACCGGATCGGGCGGTCTCGCCCGGCTTCGCGCGGACGAC 1552
Qy 1669 CCTTGGGACAGAAATTCGCGGCTTGTACGCACTCTTTCGAGCAATTCCTCCGTC 1724
Db 1553 CGCTTGGCGGATATGAAGCGGTGGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1608
```

RESULT 4

AD56825

ID AD56825 standard; cDNA; 1587 BP.

AC

AC AD56825;

XX

DT 02-DEC-2004 (first entry)

DE

XX Bacterial polynucleotide #8812.

XX

KW Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis; pathogen tolerance; pest tolerance; plant disease resistance; cell cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polynucleotide; gene; ss.

OS Bacteria.

XX

XX US2003233675-A1.

PN

XX 18-DEC-2003.

PD

XX 20-FEB-2003; 2003US-00369493.

XX

PF

XX

XX

XX

XX

XX

PR

XX

PA

PA

PA

PA

PA

XX

PI

XX

DR

XX

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Sequence 1587 BP; 261 A; 522 C; 539 G; 265 T; 0 U; 0 Other;

Query Match

13.2%; Score 244.2; DB 13; Length 1587;

Best Local Similarity 50.2%; Pred. No. 1.1e-37;

Matches 780; Conservative 0; Mismatches 743; Indels 30; Gaps 6;

Qy 178 AGCGGCGGCTGAGCTGGATCGCACATCGATCTGACACGCGCCATGCGGTGGCGAGCG 237

Db 2 ATCGGCGGACATCGACGGCCACCTGACCCCGACACCGTCCGCCCATCGCACGCG 61

Qy 238 GCGGCGGCGGATGTCCTTCCCTCCGGCGCGACCGGTGCGGTCCGTCGGAAGCGC 297

Db 62 GCCAGCGCGCGCCATCGTCCCGAGCCCGTCTCGCAAGGTTGCGGATGCCCGCGCC 121

Qy 298 GGCTCGCGGCTGTCATCGGAGCGCGCCATGCTACGACTGACACCGGCTTCGTC 357

Db 122 GTTTCGAGAGGTGGCTGCGGCCAATGTGCGGATCTACGCGGTCTCCACGGGCTTGGCG 181

Qy 358 CCCTTCGGAACCGCTGATCTCAGGTGAGATGTCCGAAGCTGCAGCGCAATCTTGTCC 417

Db 182 AACTGTACACACTGGGTGACATCGAATGCGCGTGGCTGCGTGCAGAGAACCTGTTC 241

Qy 418 ATCATCTGGCCAGCGGTGGAGCCGCTGTGATGGAACGCGCGCGGCCCATGTTTC 477

Db 242 GCAGCCATTGCGGGGTGGGTGGCTGCTGTTTTCGCGCGACGAGGTCCGCGCATGATGG 301

Qy 478 TGGCGGCTGTGGTTCGATCGCTCAGGAGGCTCCGGTGCAGCGGAGGACCATCGTC 537

Db 302 TCGCGCGTGCATGCGTTCGCGGATCTGCGCGGTGCGCGCGGTGCGCGCGGTATCGAAC 361

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.

(HINK/) HINKLE G J.

(SLAT/) SLATER S C.

(CHEN/) CHEN X.

(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 32499; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Qy 538 GCCTGATCGACCTGCTCAATTCGAGCTCGCTCCGGCCGTTCCAGCCGCGGCAACGTTGG 597
Db 362 AACTGCTGAAGTATCTGGAAGCCGGCATCAGCGCAGCCGTGCCGAGTGGTTCCGCTCG 421
Qy 598 GCGCGTGGGTGACTGACACCGCTTCGGCATATGGTGTCTGCTCCAGGCGCGGGAG 657
Db 422 GGGCCAGCGGTGATCTCGCGCTCTGCGCATCGTCCGCATCAGCTGATCGCGGAAGCA 481
Qy 658 ACTTCTCGGACCGGACGGACGGCTTGACGGCGCAGAGGCTCCGGCGCGACGGC 717
Db 482 AGTGTCTGACGAGATGCGGTACGGCACCCACGGCCGAAGTCTCTCGCGAGGTGGCA 541
Qy 718 TGCACCGCTCGATCTCTCCCATCGCGATGCACTGGCGCTGGTCAACGGGACCTTCGCCCA 777
Db 542 TCACGCGCTCGCGCTGCTACAGGAAGGGCTGGCGCTGATCAACGGGACATCGGCCCA 601
Qy 778 TGACCGGGATCGCGCTGGTGAATGCTCAGCGCTCCGCCATCTCGGCAACTGGCGGTGG 837
Db 602 TGACCGGGGTGCTGCTGCTGTGCTGAGACGCTCGCGCGCAGGTCCAGCAGGCGGAGA 661
Qy 838 CGTTGACGGCCCTGCTTTCGGAATGCTGAGAGCGCGGACCGAGGCAT---GGCGCGCG 894
Db 662 TCATCGCGGGCTGGCGCTGAAGACTATCCGCTCGGCCGATGCTTTCATGGCCCATG 721
Qy 895 CACTGTCCGACCTGCGCGCGCATCCCGGACAGAGGACGCGCAGCGAGGCTCGCGCCCC 954
Db 722 GGCACGACATCGCCAAACCGCATCCGGGACAGATCCGCTCGCGCGGCAACATGGCGCGC 781
Qy 955 GCGTGGACGGAGCGCGGGGTGTCGGGACG-----TCATTGCCGAGCGGAGGCTCG 1008
Db 782 TGCTGGCGGATTCGGGCAACGGCTCTCCGGACATGGCGAACTGTCCGCGGAGATGAAGACAC 841
Qy 1009 ACGCCGCGGATATCGGACGAGCGCGGAGCGGG-----CAGGATGCTTACAGCTGC 1062
Db 842 GCGCGGCGAGGCGCAAGAACACCGGCACTGGCGGTTCATCCAGAGGCTTACACGCTGC 901
Qy 1063 GCTGCGCTCGCGAGGTTCTCGGGCGGGCTTCGACACGCTCGCATGCAATGACCGGTGC 1122
Db 902 GCTGCATTCGCGAGTGTCTGGCGGTGGCGATACGCTCGACCATTTGGCCACCGTGG 961
Qy 1123 TGACGATCGAGCTGAACGCGGTGACGCAATCGGTGTTCCGCCCGATGGCAGCGTGC 1182
Db 962 TCGAGCGCAACTGAATTCATCGAATGACAATCGCTGTTCTTCGAAGACGGCGAGCTGT 1021
Qy 1183 CCGCCCTGACGGGGCAATTTATGGCCGAGCATGTGGCGCTGAGTCCGATCGCGCTCG 1242
Db 1022 TC-----CAGCGCGGCAACTTCACGGCCAGCAGGTGGCATTCGCAATGGACTTCTCTGG 1075
Qy 1243 CCACGGCGCTCACCGTTCGCGGGCCCTTCGGAGCGCCAGATTGCACGCTCTGACAGATG 1302
Db 1076 CCATCGCGCCACGCAACTGGCGGTGGTGTTCGGAGCGCCGCTGAACCGCTCTGAGCC 1135
Qy 1303 AAGGCTGAACCGTGGGCTCGCCCGCTTCCTCAACGGGGCCCGCGGGTTGAATTCG 1362
Db 1136 CGCATCTGAACAAACAACTGTCGGCGGTTCTGCGCGCGCGGAAACGAGGGGTTCGCTGCG 1195
Qy 1363 GCTTCATGGGGGACAGGTGACGGCGACCGGCTCTCGCGGAGATGCGAGC---CACGG 1419
Db 1196 GGTTTCGGGGGACAGATTCGGGCCACCGCGTGTGATTGCGGAGAACCGCACCATTCGA 1255
Qy 1420 GACTGCTCTCGATTCGATCTCCACGAACCGCGCAATCAGGATGTGGTCTCGCTTG 1479
Db 1256 GCCCGGAGCATCCAGAGTGTGCGCTCGAAGCGGACACACGAGATGTGTGAGCATGG 1315
Qy 1480 GGACCATCGCGCGCGCTCTGCGCGGAGAAATCGACCGTTGGCGGAGATCTTTCGGA 1539
Db 1316 GGCTGATCGCTCGCGCAATCGCGCGCATTCCTCGACAAATACCACTATCATCTCGCGC 1375
Qy 1540 TCCTCGCTCTCTGCTTGTGCAACAGCTCGGAGCTGCGCTCGGCGAGCGGCTTACGCGGG 1599
Db 1376 TGGAGTTGCTGGGTGATGTGAGCGCCCGCAACT-----CGCGGGCGCGGTGAGCAAC 1429

Qy 1600 TGTCTCCGCGGGGAAGAGCTGGTGCAGGCCCTGCGCGAGCAGTTCGCCGCCCTTGAGA 1659
Db 1430 TGGCGCGCGCAGCGCGCGCTGTTCCGCTTCGTGCGGAGCGCTGCCGCTTCTCTGCGA 1489
Qy 1660 CGACCGGCCCTGGGACAGGAATTGCCGCGCTTGTCTAGCAGCCTCTTGCAG 1712
Db 1490 TCATCGCTATATGACCGACGACATCAGGCTATGCGCGCGCTGCTCCGTCAG 1542

RESULT 5
AA63349/C
ID AA63349 standard; DNA; 42000 BP.
XX
AC AA63349;
XX
DT 06-MAR-2001 (first entry)
XX
DE Streptomyces globisporus C-1027 gene cluster ORF -7-25.
XX
KW Eneidyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;
KW cancer; ds.
XX
OS Streptomyces globisporus.
XX
FH Key Location/Qualifiers
CDS complement (8..658)
FT /*tag= a
FT /product= "ORF -7 protein"
FT complement (930..1478)
FT /*tag= b
FT /product= "ORF -6 protein"
FT complement (1649..2713)
FT /*tag= c
FT /product= "ORF -5 protein"
FT complement (2850..3237)
FT /*tag= d
FT /product= "ORF -4 protein"
FT complement (3442..4971)
FT /*tag= e
FT /product= "ORF -3 protein"
FT 5982..7479
FT /*tag= f
FT /product= "glycerol phosphate transporter"
FT complement (7573..9900)
FT /*tag= g
FT /product= "ABC transporter/UvrA-like protein"
FT complement (9982..11349)
FT /*tag= h
FT /product= "Na+/H+ transporter"
FT complement (11351..12835)
FT /*tag= i
FT /product= "hydroxylase/halogenase"
FT 13012..14079
FT /*tag= j
FT /product= "dNDP-glucose synthase"
FT complement (14212..14643)
FT /*tag= k
FT /product= "CagA"
FT complement (14690..15922)
FT /*tag= l
FT /product= "aminotransferase"
FT complement (15919..16653)
FT /*tag= m
FT /product= "N-methyl transferase"
FT complement (16653..17924)
FT /*tag= n
FT /product= "C-methyl transferase"
FT complement (18031..19191)
FT /*tag= o
FT /product= "epoxide hydase"
FT complement (19267..19929)
FT /*tag= p
FT /product= "anthranilate synthase II"
FT

Qy	944	GCTCGCGCGCGCTGTGACGGCAGCGCGCGGGTGGTCCGGGACGTCATTTCCGAGCGGAG	1003	FT	/*tag= b
Db	40189	CATCGGGCCCTGTATGGGGGCGCAGCGACTGACGGTGCAGCACGCCACCTCGCCCGAGA	40130	FT	/product= "ORF -6 protein"
Qy	1004	GCTC-----GAGCGCGCGATATCGGACGGAGCCGAGGGGGGAGGA	1048	FT	complement (1649. .2713)
Db	40129	ACTCCAGAAGACAAAGGAGCGCGCAAGGACGTCACGGCTCGAGATCTACTCTGCAGAA	40070	FT	/*tag= c
Qy	1049	TGCTACAGCTCGCGCTCGCTCCGACAGTTCTCGGGCGGGCTTCGACACGCTCGCATG	1108	FT	/product= "ORF -5 protein"
Db	40069	GGCTACTCTGCGGGCCATCCCCAGTCTGTCGGGGGGTGGCGACACCTTGATCCA	40010	FT	complement (2850. .3237)
Qy	1109	GCATGACCGGGTGTGACGATCGAGCTGAACCGCGTGACCGACAATCCGGTGTTCGCC	1168	FT	/*tag= d
Db	40009	CGCGCGCACAAGCTGGCATCGAGCTCAACTCGGCCAACGACAAACCGCTCTT-----	39956	FT	/product= "ORF -4 protein"
Qy	1169	CGATGGCAGCTGCGCCCTGCGACGGGGGCAATTTATGGGCCAGCATGTGGCGGTGAC	1228	FT	complement (3442. .4971)
Db	39955	CTTCGAGGGCAAGGAGATCTTCCACGGGGCGCACTTCCACGGTCCAGCCGATCGCGTTCG	39896	FT	/*tag= e
Qy	1229	GTCGATGCGCTCGCCACGGCGCTCACCGTTCTGGCGGGCTTCGCGAGCGCCAGATTGC	1288	FT	/product= "ORF -3 protein"
Db	39895	GATGGACTTCTGTGACATCGCGCTCACCCAGCTCGGGTCTGCGCGAGTTCCTCTCCGGGACCC	39776	FT	5982. .7479
Qy	1289	ACGCTGACAGANGAAGGTGAACCGTGGGCTGCCCCCTTCTCCACCGGGGCCCGCC	1348	FT	/*tag= f
Db	39835	CCGGGTCTGAAACCGGCACCTCAGCTACGGGCTCCCGAGTTCCTCTCCGGGACCC	39776	FT	/product= "glycerol phosphate transporter"
Qy	1349	CGGTTGAATTCGGCTTCATGGCGGCACAGGTGACGGGACCGCGTCTCTGGCCGAGAT	1408	FT	complement (7573. .9900)
Db	39775	GGGGCTGCACAGCGGATTTCGCGCGCGCCAGTACCCGGGCCACCGCACTGGTGGCCGAGAA	39716	FT	/*tag= g
Qy	1409	GCAGGCACCGGACCTCGCTCGATCCATTGATCTCCACGAACCGCCCAATCAGGATGT	1468	FT	/product= "ABC transport/UvrA-like protein"
Db	39715	CCGACATGCGCCCGCCAGCACCCAGAGGTCCCGTCCACGGCGACCAACAGGACGT	39656	FT	complement (9982. .11349)
Qy	1469	GGTCTCGTTGGGACCATCGCGCGCCCTTGC CGGAGAAATCGACCGTTGGCGGA	1528	FT	/*tag= h
Db	39555	GGTGAGCATGGCGCTGATCTCGGCGCCGCAAGCGCCCGGGTCTGTGCAACACACAA	39596	FT	/product= "Na+/H+ transporter"
Qy	1529	GATCCTTGCATCTCGCTCTGTCTTGCACAAAGCTGCGAGTGGCGTGGCGGACGG	1588	FT	complement (11351. .12835)
Db	39595	GATCTCGCG-----GTGAGTACCTTGGCGCGCCGCGCGCGTCCGACATCTCCGCGCG	39542	FT	/*tag= i
Qy	1589	CCTAGACGGGTGTCTCCCGGGGAGAAAGCTGTGTCAGGCCCTCGCGAGCAGTTCCC	1648	FT	/product= "hydroxylase/halogenase"
Db	39541	GTTGACGGCTTAGCCCCGGGGGAAAGCCACGTACGAAGCGGTGCGCGTGTGTTC	39482	FT	13012. .14079
Qy	1649	GCGCTTGAGACGGACCGCGCCCTGGGACAGGAAATTCGCGCGCTTGC	1696	FT	/product= "dNDP-glucose synthase"
Db	39481	GACGCTGGCGCTGACCGGTTACATGGCCGACGACATCGAGCTGGTCCG	39434	FT	complement (14212. .14643)
RESULT 6					/*tag= k
AAA63348/c					/product= "CagA"
ID	AAA63348 standard; DNA; 63164 BP.				complement (14690. .15922)
XX					/*tag= l
AC	AAA63348;				/product= "aminotransferase"
XX					complement (15919. .16653)
DT	06-MAR-2001 (first entry)				/*tag= m
DE	Streptomyces globisporus C-1027 gene cluster.				/product= "N-methyl transferase"
XX					complement (16653. .17924)
KW	Eneidine C-1027 biosynthesis gene cluster; apoprotein; chromophore;				/*tag= n
XX	cancer; ds.				/product= "C-methyl transferase"
OS	Streptomyces globisporus.				complement (18031. .19191)
PH	Location/Qualifiers				/*tag= o
Key	complement (8. .658)				/product= "epoxide hydrazase"
FT	/*tag= a				complement (19267. .19929)
FT	/product= "ORF -7 protein"				/*tag= p
FT	complement (930. .1478)				/product= "anthranilate synthase II"
FT					complement (19926. .21407)
FT					/*tag= q
FT					/product= "anthranilate synthase I"
FT					complement (21424. .22878)
FT					/*tag= r
FT					/product= "coenzyme F390 synthetase"
FT					complement (22875. .23546)
FT					/*tag= s
FT					/product= "iron-sulphur flavoprotein"
FT					complement (23566. .24702)
FT					/*tag= t
FT					/product= "O-acyl transferase"
FT					complement (24986. .25564)
FT					/*tag= u
FT					/product= "epimerase"
FT					25815. .27170
FT					/*tag= v
FT					/product= "monooxygenase"
FT					27214. .28593
FT					/*tag= w
FT					/product= "glycosyl transferase"
FT					28590. .29588
FT					/*tag= x
FT					/product= "dNDP glucose dehydratase"
FT					29632. .31197
FT					/*tag= y
FT					/product= "SgcB transmembrane efflux protein"
FT					31280. .32590
FT					/*tag= z

Qy	707	CGCGGACGCGTGCACACCGCTCGATCTCTCCCATCCGATCCGATCGACATGCGCGCTGGTCAACGG	766
Db	40429	CGACGCGGGGCATCGAGCCGCTCGAACTGCGCTTCAAGGAGGGCCCTCGCATGATCAACGG	40370
Qy	767	GACCTCCGCGCATGACCGGGATCGCGTGTGTGAATGCTCACGCGCTCGCGCATCTCGGCAA	826
Db	40369	CAGTCCGGGATGACCGGTCTGGGCTCCCTGTGCTCGACGGGCCCTTGGACGAGGCCCA	40310
Qy	827	CTGGCGGGTGGCGTTTACCGGCCCTGTCTTGGGAATGTCTGAGAGCCCGGAACCGAGGCATG	886
Db	40309	GCAGGCCGAGATCGTGAACGCGCTCTGCTCATCGAGGCGGTACCGCGATCGACCGAGCCCTT	40250
Qy	887	GGCGCGGCGACTGTCCGACTCT---GGGCGCGCATCCCGACAGAGGACGCGCGAGCGAG	943
Db	40249	CCTCGCGGAGGGGCAAGACATAGCCCGCCGACGAGAGGCCAGATCGAACCGCGGCCCAA	40190
Qy	944	GCTCGCGCCCGCGGTGGACCGGACGCGCGGGGTGGTCCGGGACGTCATTTCCGACGCGAG	1003
Db	40189	CATGCGGGCCCTGATGCGGGGACGCGACTGACGGTTCGAGCACGCCGACCTGCGCCGAGA	40130
Qy	1004	GCTC-----GACGCCGGCGNATTCGGGACGGAGCCGGAGGCGGGGACGAGA	1048
Db	40129	ACTCCGAAGGGAACAAGGAGGCCCGGCAAGGACGCTCCAGCGCTCGGAGATCTACTCTCGAA	40070
Qy	1049	TGCTACAGCCTCGCTCGGCTCCGCAAGTTCCTCGGGCGGGCTTCGACACGCTCGCATG	1108
Db	40069	GGCTTACTCGCTCGGGCCATCCCCCAGGTCTGTGGGGGGGTGCGGACACCTTGTACCA	40010
Qy	1109	GCATGACCGGGTCTGACGATCGAGCTGAACCGGTGACCGGCAATCTCGGTGTTCCTCGCC	1168
Db	40009	CGCGGGSACAAGCTGGCATCGAGCTCACTCGGCCAACGACACCCGCTCTT-----	39956
Qy	1169	CGATGGCAGGCTGCCCGCCCTGACCGGGGGCAATTTCTAGGGCCACGATGTGGCGCTGAC	1228
Db	39955	CTTCGAGGGCAAGGAGATCTTCCACGGGGCGAACTTCCACGGTCAAGCCGATCGCGTTGCG	39896
Qy	1229	GTCCGATGCGCTCGCACGCGCTCACCGTCTTGGCGGCGCTTCGGGAGCGCCAGATTGC	1288
Db	39895	GATGGACTTCGTGACCATCGCGCTCACCCAGCTCGCGGTCTGGCGGAGCGGAGATCAA	39836
Qy	1289	ACGTCTGACAGATGAAGGCTGAACGCTGGGCTGCCCCCTTCTCCACCGGGGCCCCGC	1348
Db	39835	CCGGTCTCTGAAACCGGCACTCAGTCAAGGCTCCCGGAGTTCTCTGCTCCGGGGACCC	39776
Qy	1349	CGGGTTGAATTCGGGCTTCATGGGCGCCACAGGTGACGGCGACCGCGCTCTGGCGCGAGAT	1408
Db	39775	GGGGCTGCAACAGCGGATTCGCGGCGCCAGTACCGCGGCCACCGCACCTGGTGGCGCGAGA	39716
Qy	1409	GCAGGCCACGGGACCTGCGCTCGATTCGATCTCTCCAGAACGCGGCCAATTCAGGATGT	1468
Db	39715	CCGGACGATCGGCGCGCCAGCACCCAGAGCGTCCGCTCCACGCGGGAACACCGAGAGCT	39656
Qy	1469	GGTCTCGTTGGACCATCGCGCGCGCTCTTCCGCGGAGAGATCGACCGTTGGGCGGA	1528
Db	39655	GGTGAGCATGGGCGTGTATCTCGGCGCGCAACGCCCGCGGGTCTGTGGAACAACAA	39596
Qy	1529	GATCCTTGGCATCCTCGCTCTGTGCTTTCGACAGCTCGGGAGCTCGCTGCGGCGAGGG	1588
Db	39595	GATCCTCGCG-----GTGGAGTACTGTGCGCGCGCCACGGCGGTTCGACATCTCGGCGG	39542
Qy	1589	CCTAGACGGGGTGTCTCCCGCGGGGAAGAGCTGGTGTGAGGCGCCTGCGCGAGCAGTTCCC	1648
Db	39541	GTTTCGACGGCTTGAGCCCGGCGGGAAGGCCACGTACGAGGCGGTGCGCGGCTGTTCC	39482
Qy	1649	GCGCGTTGAGACGGACCGGCCCTTGGGACAGGAATTTGCCGCGCTTGC	1696
Db	39481	GAGCTTGGGCGTGCAGCGGTACATGTCGCGGACGACATCGAGCTGGTTCG	39434

RESULT 7
ADS55996
ID ADS
XX

AC	AUS55996;
XX	
XX	02-DEC-2004 (first entry)
XX	
XX	Bacterial polynucleotide #7983.
XX	
XX	Recombinant DNA construct; transformed plant; improved plant property;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW	pathogen tolerance; pest tolerance; plant disease resistance;
KW	cell cycle pathway modification; plant growth regulator;
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW	bacterial polynucleotide; gene; ss.
XX	
OS	Bacteria.
XX	
XX	US2003233675-A1.
XX	
XX	18-DEC-2003.
XX	
XX	20-FEB-2003; 2003US-00369493.
XX	
XX	21-FEB-2002; 2002US-0360039P.
XX	
XX	(CAOY/) CAO Y.
PA	(HINK/) HINKLE G J.
PA	(SLAT/) SLATER S C.
PA	(CHEN/) CHEN X.
PA	(GOLD/) GOLDMAN B S.
XX	
XX	Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX	
DR	WPI; 2004-061375/06.
XX	
XX	New recombinant DNA construct comprising a promoter positioned to provide
PT	for expression of a polynucleotide encoding a polypeptide from a
PT	microbial source, useful for producing plants with improved properties.
XX	
XX	Claim 1; SEQ ID NO 31670; 122pp; English.
XX	
CC	The invention relates to a recombinant DNA construct comprising a
CC	promoter functional in a plant cell, where the promoter is positioned to
CC	provide for expression of a polynucleotide encoding a polypeptide from a
CC	microbial source. The invention also relates to a transformed plant
CC	comprising the recombinant DNA construct and a method of producing a
CC	transformed plant having an improved property. The plant is a crop plant
CC	such as maize or soybean. The method of producing a transformed plant
CC	having an improved property comprises transforming a plant with the
CC	recombinant DNA construct and growing the transformed plant, where the
CC	polynucleotide or polypeptide is useful for improving plant properties.
CC	The recombinant DNA construct is useful for producing plants with
CC	improved plant properties, e.g. improved cold, heat or drought tolerance,
CC	tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC	increased resistance to plant disease, better growth rate by modification
CC	of the cell cycle pathway with plant growth regulators, increased rate of
CC	homologous recombination, modified seed oil or protein yield and/or
CC	content, improved yield by modification of carbohydrate, nitrogen or
CC	phosphorus use and/or uptake, by modification of photosynthesis or by
CC	providing improved plant growth and development under at least one stress
CC	condition. Improved lignin production or improved galactomannan
CC	production. This sequence represents a bacterial polynucleotide used in
CC	the scope of the invention. Note: The sequence data for this patent did
CC	not form part of the printed specification but was obtained in electronic
CC	format from USPTO at seqdata.uspto.gov/sequence.html .
XX	
SO	Sequence 1497 BP: 191 A: 577 C: 527 G: 202 T: 0 U: 0 Other;

Query Match 12.9%; Score 238.4; DB 13; Length 1497;
Best Local Similarity 50.8%; Pred. No. 1.4e-36;
Matches 73; Conservative 0; Mismatches 666; Indels 51; Gaps 5

Qy 199 GCACATCGATCTGGACCAGGCCCATCCGTGCGAGCGGC CGCGGATTGTCCTTG 258
||||| | |||| | |||| | |||| | |||| |
| | |

Db 14 GCACAGCCAGCTCGCGCAGCTCGAAGCGATCTGGCGAGGGGGCGCGCGCTCTGG 73
Qy 259 CCCCTCCGCGCGGACCGGTGCTGCTCGAAGCGGGGTGCGCGCTGTCATCCCGC 318
Db 74 CCCCCGAGCGCGCGCCCGCTCGAGAGGGCGCGCGCGCTGGCCGAGGCGCGCGC 133
Qy 319 AGGCGCGCATGCTACGAGACTGACAAACCGCTTCGGTCCCTTGGCGAAGCGCTGATCT 378
Db 134 GCACGCGCGCTCTATGCGGTGAACACGCGCTTCGGCAAGCTCGGAGCTCTAAGATCG 193
Qy 379 CAGGTGAGAAATGTCGAAACGCTGCAAGGCAATCTTTGTCATCATCTGGCCAGCGCGTGG 438
Db 194 CTCGCGCGCATACGCGCGCAACTGCAACGCACTGATCTGTGCACTGCTGGCGGTG 253
Qy 439 GACCGGTCTTGAATGAGAGCGCGCGCGCATGCTTCTGGCGGTCTGTTGTCGATCG 498
Db 254 GCGAGCTATGCCCCCGCTCCAGCGCGCGCTGATGATTCGCTGAAGCTCTGTCGCTCG 313
Qy 499 CTCAGGAGACCTCCGCTGCCAGCGAGGGGACCATCGCTCGCTGATCGACCTGCTCAATT 558
Db 314 CCGCGCGCTTCGGCGTGGCTGGAGATGTTGGCGCTGCTCGAAGGCAATGCTGGCGC 373
Qy 559 CCGAGCTCGCTCCGCGCTTCCAGCCCGCGACGGTGGCGCGCTGCGGTGACCTGACAC 618
Db 374 CCGCGCTCACGCGCGTATCCGCGCGCAGGGGTGGTCCGCGCGAGCGCGATCTGGCAC 433
Qy 619 CGCTTGGCATATGTTGCTCTGCTCCAGCGCGGGGAGACTTCTTGGACCGGGACGGGA 678
Db 434 CCGTGGCCCATATGGCCCGCTGATGATCGCGAGGGCGAGCGGAGGTC---GGCGGCA 490
Qy 679 CGCGCTTGGCGCGACGAAGGGTCCGCGCGGAGCGCTGCAACCGCTCGATCTCTCC 738
Db 491 GCGCTGCTCCGCTGCGCGCTGGCGCTGGCGAGCGGCTTGGCCCGGTGGCCCTCGGAC 550
Qy 739 ATCGCATGCACTGCGCTGCTGCTCAACGGGAATCTCGCCATGACCGGGATCGCGTGTGA 798
Db 551 CCAAGGAGGGCTCGCCCTCATCAACGGCACGCAATCTCGACCGCTATGCCCTCGCGC 610
Qy 799 ATGCTACCGCTCGCGCATCTCGGCAACTGCGCGGTGGCGGTGACGGCCCTGCTTGGG 858
Db 611 GCCTCTTCGAGGGTGGCGCGCTCGAGGGCGCTGATCTCGGCGCTCTCCACCG 670
Qy 859 AATGCTGAGAGGCGGACGAGCATGGCGCGGACACTGTCGACCTGGCGCGCATC 918
Db 671 ATGCGATATGGGTGCACTGCGCGCTCCCGCGGATCCAGCGCTGCGCGGCATG 730
Qy 919 CCGGACAGAGGACCGCGACGAGGCTCGCGCGCGCGCTGAGCGCGCGCGGTGG 978
Db 731 CGGCGCAGATCGAGGGCGCGCCACCATGCGCGCTGCTCGAAGGCTCGGCCATCGCG 790
Qy 979 TCCGCGACGTCAITTCGAGCGGAGGCTCGACGCGCGGATATCGGAGCGGAGCGGAGG 1038
Db 791 AGAGCCACCGTAGGGGACCC-----AGC 814
Qy 1039 CGGCGCAGGATGCTACAGCTCGCTCGCTCGCGAGGTTCTCGGGCGGCTTCGACA 1098
Db 815 GGGTGCAGGACCCCTACTGCACTCGCTGCGAGCGCGAGGTGACGGGCGCGCGATGATG 874
Qy 1099 CGCTCGCATGTCATGACCGGTGCTGACGATCGAGCTGAACCGCGGTGACCGCAATCCG 1158
Db 875 TGCTCGCATGCGCGCGGCGAGCTGGCCACCGAGGCAATGCGCCACCGCAATCCG 934
Qy 1159 TGTTTCGCGCGATGCGAGCGTGCCTCGCTGCAAGGGGCAATTTATGGCGCAGATG 1218
Db 935 TCGTGTCTTCGGA-----CGGCGCATGCTGCTCGGAGGCACTTCATGCGGAGCCG 988
Qy 1219 TGGCGCTGACCTCGATGCGCTGCCACGCGCGCTCACCGTTCTGCGGGCTTCGGAGC 1278
Db 989 TGGCTTCGCGCGCGACATGATCGCTGCGCGCTCTCGAGATCGGCGCGCATCGCGAGC 1048
Qy 1279 GCCAGATTGCACTGTCAGACATGAAGGCTGAACCGGTGGCGCTGCCCGCTCTCTCCACC 1338
Db 1049 CCGCGTGGCGCTGATGTTGATCGACGCTCTCTCTCGACCTTCGCGCTCTCTCACCC 1108

RESULT 8

ACA25972
ID ACA25972 standard; DNA; 1521 BP.
XX ACA25972;
AC ACA25972;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #7629.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
OS Burkholderia mallei.
PN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 23-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trarwick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.
P-PSDB; ABU22102.

New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 13842; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated

polypeptide or its fragment whose expression is inhibited by the
antisenese nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisenese nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is one of the target
prokaryotic essential genes. Note: The sequence data for this patent did
not form part of the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

Sequence 1521 BP; 220 A; 548 G; 530 G; 223 T; 0 U; 0 Other;

Query Match 12.5%; Score 231.6; DB 8; Length 1521;
Best Local Similarity 50.9%; Pred. No. 2.8e-35;
Matches 753; Conservative 0; Mismatches 664; Indels 63; Gaps 6;

229 TGGCGAGCGGCGCGGATGCTCTCCCTCCGCGCGCGAGCGGTCGCGT 288
50 TCGCCCCGAGAACGTGCAGATCGCTCGATCCCGGAGCTTCGCGCGATCGACCGG 109
289 CCGAAGCGCGGCTCGCGCTGTATCCGCGAGGCGCCATGTCTACGACTGACAAACCG 348
110 GCGCGAGGCGCTCGCGACATCCCGAAGGCGGAGCGGCTACGGCATCAACACGG 169
349 GCTTCGCTCCCTTGGCAAACCGCTGATCTCAGGTGAGATGTCCGAACCGCTGAGGCCA 408
170 GCTTCGCGCGCTCGCGAGCAGCACATTCGACGACGACGCTCGAGCTCTGACAGA 229
409 ATCTGTGTCATATCTGCCAGCGGCTGGACCGGTCTGACTGACACGACGCGCGG 468
230 ACTTCGTGCTGTGACGCGGTGGGCTCGCGAGCGCGATGGCGCGCCCGTCTGTCGCC 289
469 CCATGGTTCCTGGCGCTGCTGTGTCATCGCTCAGGGAGCGCTCGGTGCGCAGCGAGGGA 528
290 TGTGTATGGCGCTCAAGCTCTGAGCTTCGCGCGGCGACTCGGGCATTCGTGCGGTG 349
529 CCATGCTCGCTGATCGACCTGCTCAATTCGAGTCTCGTCCGCGCGGTTCCAGCGCG 588
350 TGATGGACGCGCTGCTCACTGCTGTTCAACGCGGACGCTGCTGCGCTCATTCGCTCAAG 409
589 GCACGTTGGGCGGCTCGGGTGACTGACACCGCTTGGCGATATGGTCTCTGCTCCAGG 648
410 GCTCGGTGGGCGGCTCGGGGCACTTCGCGCGCTCGCGCGCATGTGCGCGCTGCTGCTG 469
649 GCGCGGAGACTTCTGGACGCGGACGCGGCTTGACGCGCGAGAGGCTCCGGC 708
470 GCATCGGCGAGCTGTTTCATCGGGCA---GCGCGGAGCGGCGCGAAGGACTGCGTG 526
709 GCGGACGGCTGCAACCGCTCGATCTTCCCATCGCATGCACTGGCGCTGTTCAACGGGA 768
527 TCGCGGCGCTCGCGCGCTCACTGCTCAAGCGAAGGAGGCGCTCGCGCTGCTGAACGGCA 586
769 CTTCCGCGATGACCGGATCGCTGTGTAATGCTACGCGCTGCGCGCATCTCGGCACT 828
587 GCGAGGCGTGCAGCGCGCTCGGCTGCACAACTGTTTCGCGATCGAGGACCTGTACCGCA 646
829 GGGCGGTGGCTTGACGGCCCTGCTTCCGGAATGCTCTGAGAGGCGCGACCGAGCATGGG 888

647 CGCGCTCGTGTGCGGCGCGCTGCTCGGTGCACGCGCGCGCGCTCGGTGAAGCGTTCG 706
889 CCGCGGCACTGTCCGACCTTGGCGCGCATCCCGGACAGAAAGGACGCGCAGGAGGTGC 948
707 ACGGCGCATCCACAGAGCTCGCGGCGCATCGCGCCAGATCGACGCGCGCGCGTACC 766
949 GCGCGCGCTGGACGCGCGCGGCTGTGTCGCGCACGTCATTGCGGAGCGGAGGTTCG 1008
767 GGTGCTGCTCGAGCGCTCGGCGATCAACGTGTGCGACCGC----- 807
1009 ACCCGCGCATATCGGGACGAGCCGCGGCGGAGGATGCTACAGCTCGGCTGCG 1068
808 -----GATTGCGAAGGTGAGGACCGCTACAGCTCGCTGCC 847
1069 CTCGCGAGTTCGCGGCGGCTTCGACACGCTCGCATGGCATGACCGGCTCTGACGA 1128
848 AGCGCGAGTGTGGGCGGCTGTCTCGACAGATCCGCCACGCGCGCGGCTGTGCTCA 907
1129 TCGAGCTGAACGCGGTGACCGACAATCCGCTGTTCCGCCGATGGCAGCGTCCCGCC 1188
908 TCGAGCGGAACGCGGTGTGGACAAACCGCTGATCTCCCGGACACGGGCGAG---GTGC 964
1189 TGCACGCGGCGCAATTCATGGGCGCAGCATGTGGCTGACGTCGATGCGCTCGCACGG 1248
965 TGTGCGGCGCAATTCACGCGGAGCGCTCGCTTCGCGCGCGACAAATCTCGCGATCG 1024
1249 CGGTACACGTTCTGGCGGCGCTTCGCGAGGCGCAGATTGACGCTCTGACAGATGAAGGC 1308
1025 CCGCGCGGAGATCGCGCGCTCGCGAGGCGCGCATCGCGTGTGATCGACGCGACGC 1084
1309 TGAACGCTGGGCTGCGCGCTTCTCCACGCGGCGCGCGGTTGAAATTCGCGCTTCA 1368
1085 TCTCGCGCTGCGCGC-----TTTCTGTGAGAGCGCGCGCTGAACTCGGCTTCA 1138
1369 TGGCGCACAGGTGACGCGGACCGCGCTCTTGGCGAGATGCGAGC---CACGGGACCTG 1425
1139 TGATCGCGCAGGTGACGCGCGCGCTCGCGTCGGAAACAAAGACGCTCGCGCATCGG 1198
1426 CCTCGATCATTCGATCTCCAGAACCGCGCCATCAGGATGGTCTCGCTTGGGACCA 1485
1199 CGTGGTCAATTCGCTGCGGAGCTCGCGAACCGAGAACCAACGTCGATGCGGACGT 1258
1486 TCGCGCGCGCTCTCGCGGAGAGATCGACGTTGGGCGGAGATCCTTGCGATCTCG 1545
1259 TCGCGCGCGGACGCTCAGGACATCGCGAGAACGTCGGGAACATCTCTCGCATCGAGC 1318
1546 CTCTCTGTCTGCACAAGCTGCGGAGCTGCGTTCGCGGAGCGGCTTAGACGGGTTGTC 1505
1319 TGCTCGCGCGCGCAAGCGCTCGACCTGC-----GCGCGCGCACGCAACGAGCC 1369
1606 CCGCGGGGAGAGCTGTGTCAGGCGCTCGCGAGCAGTTCCCGCGCTTGAGACGAGCC 1665
1370 CGCGCTGCGACGCGATGAAGACGATTCGCGCGGACGCTCGCGCACTACGATCTCGACC 1429
1666 GCGCGCTGCGACAGGAATTCGCGCTGTCACGCACT 1705
1430 ACTACTTCGCGCGCGCATCGCGTGTGTCGCGCGCGCT 1469

RESULT 9

ADT44460

ID ADT44460 standard; cDNA; 1503 BP.

XX ADT44460;

XX 02-DEC-2004 (first entry)

XX Bacterial polynucleotide #19211.

XX Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;

cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 42898; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1503 BP; 225 A; 535 C; 516 G; 227 T; 0 U; 0 Other;
Query Match 11.6%; Score 215.4; DB 13; Length 1503;
Best Local Similarity 50.0%; Pred. No. 3.7e-32;
Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
203 CATCGATCTGACACGAGCCCATGCGGTGGCGAGCGCGCGCGGAGTGTCTTGGCCCC 262
21 CCTGAAGCTCGAGGAATCTCCAGTGGCTCGCAACGAGGCCACCGTGGAGCTGTGCC 80
263 TCCGGCGCGGACCGGTGCGGTGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 322
81 CGAGCGCCGCCACCGCGGTGCGGGGCTCGCGCGGCTGTGGTGGTGGTGGTGGTGGTGG 140
323 GCGGCATGCTACCGACTGCAACCGGCTTCGGTCCCTTCGGTGGTGGTGGTGGTGGTGG 382
141 CACGCCCGCCCTACGGCATCAACACCGGCTTTGGCACGCTTGGCGAGGTTCGCGCATCG 200

QY 383 TGAGATGTCCGAACGCTGACGGCCAAATCTTGTTCATCATCTGGCCACGCGCGTGGACC 442
DB 201 GAAGGACCTGCGGACCTCCAGCGCAACCTATCTCTCCACACGCTGTGCGTGGCAC 260
QY 443 GGTGCTTGTGACGTGACGCGCGCGCATGTGTTCTGSCGCTCTGTGTGTGATCGCTCA 502
DB 261 GCCCTTCCCTTCCGGAAGCGCGGCGCTCTGTGTCTGCTGCAACGTGTCTCGCCAA 320
QY 503 GGGAGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
DB 321 GGGCTACTCTCGGCGATCCGATGAGAGCGTGGCCCTGSCGCTGAGCATGTCTGNAAC 380
QY 563 GCTGCTCCGCGCGTTCAGCGCGCGGACGCTGCGCGCGCGCGCGCGCGCGCGCGCG 622
DB 381 CGTGTGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 440
QY 623 TGGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
DB 441 GCGCACTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 497
QY 683 GCTTGAAGCGCGAGAGGCTTCCGCGCGCGGAGCGCTGCAACCGCTGCTGCTGCTGCT 742
DB 498 GATGCCCGGAGAGCAGCGCTGAGCGCGCGCGCGCTGCAACCGGTGCTGCTGCTGCTG 557
QY 743 CGATGCACTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802
DB 558 GGAGGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 617
QY 803 TCACGCTGCGCGCATCTCGGCAACTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 862
DB 618 GCTTTCGCGGAGTCTCTGCGGACATGCGCGCGCGCGCTGCAACCGGTGCTGCTGCTG 677
QY 863 TCTGAGAGCGCGAGCCGAGCGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 922
DB 678 GCTGCTGGAAGCCCAAGCCCTTCTTCTGAGATTCAACGCTGCGCGCGCGCGCGCG 737
QY 923 ACAGAGAGCG 982
DB 738 CCAGAGAGCGCTGCG 797
QY 983 GCAGCTCATTTCCGAGCGGAGGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1042
DB 798 GCAGCTCA-----ACTGCAAGAGT 818
QY 1043 GCAGATGCTCTACAGCTGCGCTGCGCTGCGCGCGCGCGCTGCGCGCGCGCGCTTCC 1102
DB 819 GCAGGACCCCTACTCTCTGCGCTGCTGCGCGCGCGCGCTGCAACCGCGCGCGCGCG 878
QY 1103 CGCATGGCATGACCGGCTGCTGAGCTGAGCTGAAACGCGGTGACGACAAATCCGCTGT 1162
DB 879 CGCGTCTCTCCGCGCGCATCTCTGAGGTGAGGTCAACAGCGCGCGCGCGCGCGCGCT 938
QY 1163 TCCGCCGATGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1222
DB 939 ---GTTCCGCGACACGAGGCGCATGCTGCTGCGCGCGCGCAACTTCCACGCGCGCG 995
QY 1223 GCTGAGCTCGATGCGCTCGCCACGCGCGCGCTGCAACCGCTGCGCGCGCGCTTGGCG 1282
DB 996 CTTGGCCATGAGCGTGTGGCGATGGCGCTGACGCAACTGTCTGCTCATCAGCGAGCG 1055
QY 1283 GATTGCACTGTGACATGAAAGGTGAAACCGGTGAGCGCGCGCGCGCGCGCGCGCG 1342
DB 1056 CGTGGAGCAGCTGTGAAACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1109
QY 1343 CCGCGCGCGGTGAAATTCGCGCTTCAATGGGCGCACAGTGAACGCGCGCGCGCGCGCG 1402
DB 1110 GAACTCCGCGGTGAACTTCGCGCTTCAATGATCGCGCGCGTGAACGCGCGCGCG 1169
QY 1403 CGAGATCGGAGCGCGGGA---CCTGCTCGATCCATTTCGATCTCCAGAACGCGCGCAA 1459
DB 1170 CGAGTCCCGCTGCTGAGCGCACCCCGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTG 1229

QY 1460 TCAGATGTGTCTCGCTTGGGACCATCGCGCGCGCTCTCGCGCGAGAGATCGACCG 1519
DB |||||
1230 AGAGACCATCGTGTCCATGGGCATGACGGCGCGCTCAAGGGCGCTCAGGTTCAGCGACTT 1289
QY 1520 TTGGGCGGAGATCCTTGGATCCCTCGCTCTCTGTCTTTCACACAGCTGGGAGCTGGCGTG 1579
DB |||||
1290 CGCCGCTTGTGCTCGCATTTGAAATTCCTGTGTGGCGCGGAGGCCCTGGACTTCGCGCT 1349
QY 1580 CGGACGGCGCTAGACGGGGTGTCTCCCGGGGGAAGAAGCTGTGTGACGGCCCTGCGCGA 1639
DB |||||
1350 GCCGCTGAAGCCCGCAAGGGCGCTTGGGCGTACGAGCTGG-----TGCGCTC 1400
QY 1640 GCAGTCCCGCGCTTGGAGCGACCGGCCCTTGGGACAGAAATTCGCGCGCTTGTCTAC 1699
DB |||||
1401 GAAGTCCCCACATGACAGGACCGGAGCTGCACCGGACATCGAGGCGGTGAGCCA 1460
QY 1700 GCACCTC 1706
DB |||||
1461 GCTCGTC 1467

RESULT 10
ACL71540
ID ACL71540 standard; DNA; 1527 BP.
XX AC
XX ACL71540;
XX 02-JUN-2005 (first entry)
XX M. xanthus gene sequence, seq id 8003.
XX Transgenic plant; DNA replication; gene regulation; gene expression;
XX gene; ds.
XX Myxococcus xanthus.
XX US6833447-B1.
XX 21-DEC-2004.
XX 10-JUL-2001; 2001US-00902540.
XX 10-JUL-2000; 2000US-0217883P.
XX (MONS) MONSANTO TECHNOLOGY LLC.
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX WPI; 2005-028716/03.
XX New substantially purified Myxococcus xanthus nucleic acid molecule
XX encoding a nitrite reductase, useful for determining gene expression,
XX identifying mutations in a gene of interest, and for constructing
XX mutations in a gene of interest.
XX Example 2; SEQ ID NO 8003; 25pp; English.
XX The invention relates to a substantially purified nucleic acid molecule
XX encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
XX recombinant DNA construct for expression of a nitrite reductase gene in a
XX plant cell, and a plant cell comprising the recombinant DNA construct.
XX The nucleic acid is useful for determining gene expression, identifying
XX mutations in a gene of interest, and for constructing mutations in a gene
XX of interest. Sequences given in records for SEQ IDs 1850-9691 represent a
XX set of about 7842 genes or partial genes from the genome of the bacterium
XX Myxococcus xanthus. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from USPTO
SQ Sequence 1527 BP; 228 A; 545 C; 523 G; 231 T; 0 U; 0 Other;

Query Match 11.6%; Score 215.4; DB 14; Length 1527;
Best Local Similarity 50.0%; Pred. No. 3.7e-32;

Matches 753; Conservative 0; Mismatches 691; Indels 63; Gaps 6;
QY 203 CATCGATCTGGACACAGCCCATCGCGCTGGCGAGCGCGCGCGATTCCTTGGCCCC 262
DB |||||
36 CCTGAAGCTCGAGGAATCTCCAGGTGGTCCGAACGAGGCCACCGTGAAGCTGTGCC 95
QY 263 TCCGGCGCGACACCGGTGCGTCCGAAGCGCGCTCGGCGCTGTCTATCCGCGAGGC 322
DB |||||
96 CGAGCGCGCACCGCGTGGCGCTCGCGCCCTGGTGGACCGGTGCCGCCGAGA 155
QY 323 GCGCATGTCTACGAGTGAACACCGGCTTCGGTCCCTTGGAAACCGCTGATCTCAGG 382
DB |||||
156 CAGCCCGCGCTACGGCATCAACACCGGCTTGGCAGCTTGGCGAGTCCGCATCGACAA 215
QY 383 TGAGATGTCCGAACGCTGAGGCCCAATCTGTTCATCATCTGCGCAGCGCGGTGGGACC 442
DB |||||
216 GAAGGACCTGCGGACCTCGAGCGCAACCTATCTCTCCACGCGTGTGGCTGGGCAC 275
QY 443 GGTGCTTGACTGGAACGCGCGCCCATGGTTCCTGGCGCGTCTGGTGTCTGATCGTCA 502
DB |||||
276 GCCCTTCCCTTCCGGAAGCGCGGCGCTCTGCTGCTCCGCTGCAACGTGTGGCAA 335
QY 503 GCGAGCTCTCGGTGCGAGGAGGGGACCATCGCTCGCTGATCGACTGCTCAATTCGA 562
DB |||||
336 GGGCTACTCGGCATCGCATGGAGACGCTGGGCCCTTGGCGCTGGACATGCTGAACCGGGA 395
QY 563 GCTCGCTCCGGCGCTTCCAGCGCGCGCGGCGGTGGGCTGCGGTGACCTGACACCGCT 622
DB |||||
396 GCTGTGCGCGTGTCTCCGAGCGGGGCGAGTGGGCGCTCCGGGATCTCGCCCGCT 455
QY 623 TCGCATATGTGTCTCTGCTTCCAGGGCCGGGAGACTTCTCTGGAACGGGACGGGACGG 682
DB |||||
456 GCGCACCTGGGCTCTGCTTTCATCGCGAAGGTGAAGCTTCTATCAGGGCCA---GCG 512
QY 683 GCTTGAAGGGGAGAGGGCTCCGGCGGACCGGTGCAACCGCTGATCTCTCCATCG 742
DB |||||
513 GATGCCCGGAGACGAGCGCTGGAGCGCGCGCGCTGCAACCGGTGTGTTGAGGCGCAA 572
QY 743 CGATGCACTGGGCTGTGTAACGGGACCTCCGCCATGACGGGATCGCGTGTGTAATGC 802
DB |||||
573 GGAGGGCTCGCTCTGTGAACCGGACACAGGCATGTGGCGGTGGGACACCTGTCTCCA 632
QY 803 TCACGCTTCCGCGCATCTCGGCAACTGGGCGGTGGCGTTGACGGCGCTGTCTCGGAATG 862
DB |||||
633 GCTTCGCGGGAGTCCCTTGGCGGACATGCGCGAGTCCGCGGGGCCATGACGCTGAGGG 692
QY 863 TGTGAGAGCGGACCGGAGCATGGGCGCGGCACTGTTCGACCTGCGGCGCATCCCGG 922
DB |||||
693 GCTGTGGGGAAGCCACAAGCCCTTCAATCTGAGATTCAACGATTCACGACGTCGCGCGCACCGGG 752
QY 923 ACAGAGGACCGCGCAGCGGCTGCGCGCGCGTGGACGGCAGCGCGGGTGTGCG 982
DB |||||
753 CCAGAGGACGTCGCGGCGACCTCGGCGCATCTCTGGTGGACAGCGAGCTGTGTGAGTC 812
QY 983 GCACGTCTATTGTCGAGCGGAGGCTCGACGCGCGCGCATATCGGGACGGAGCCGAGGCGGG 1042
DB |||||
813 GCACGTCA-----ACTGACGAAGGT 833
QY 1043 GAGGATGCTCAGCGCTGGCTGCGCTCGCAGGTTCGCGGGCGGGCTTCGACAGCT 1102
DB |||||
834 GCAGGACCCCTACTCTCGCTGCAATGCGGAGGTGACGCGCGCGCGCGCGGAGGCGCAT 893
QY 1103 CGCATGCGCATGACCGGGTGTGACGATCGAGCTGGAACGCGGTGACACGAATCCGGTGT 1162
DB |||||
894 CGGTTCTCCCGCGCATCTCGAGGTGGAGGTCAACAGGCGACGAGCAACCCGCTCGT 953
QY 1163 TCCGCCCGATGGCAGCGTCCCGCTTCGACCGGGGCAATTTTCATGGCGCAGCATGTGGC 1222
DB |||||
954 ---GTTGCGGACACGAGGGCATCTGTGTCGGGCGCACTTCCACGCGCAGCCCATCTC 1010
QY 1223 GCTGAGCTCGATGCGCTCGCACGCGCGCTCAACGTTCTGGCGGGGCTTGGCGAGGCCA 1282
DB |||||
1011 CCTGGCCATGGACGTGTGTGGCGATGCGCGCTGACGCAACTGTCTCATCAGCGCGCGG 1070


```
Db 2153 GCAGGACCCCTACTCCCTGCGCTGCATGCGCAGGTGCACGGCGCGCGCGAGGGGCAT 2212
QY 1103 CGCATGCGATGACCGGGTGTGACGATCGAGCTGAACGCGGTGACGACAAATCCGGTGT 1162
Db 2213 CGGTCTCCGGGGCATCTCGAGGTGGAGGTCAACAGCGCGACGGAACCCGCTCGT 2272
QY 1163 TCCGCCCGATGGCAGCGTGCCTCCCTGCAACGCGGGCAATTTTCATGGCCGAGCATGTGC 1222
Db 2273 ---GTTCCGGGACACGAGGCGCATCGTGTGCGGCGCAACTTCCAGCGCCAGCCCATCTC 2329
QY 1223 GCTGAGCTCGATGCGCTCGCCAGCGCGCTGACCGTCTGCGGGGCTTCGAGGAGGCCA 1282
Db 2330 CCTGGCCATGAGCGTGGTGGCGATGCGGCTGACGCAACTGTCTGCTCATCAGCAGCGCG 2389
QY 1283 GATTGCACTGTCACAGATGAAGGCTGAACCGTGGGCTGCCCCCTTCCTCCACCGGG 1342
Db 2390 CGTGGAGCAGCTGTGNAACCGCTGCTGTCCAACTGCGGGCT-----TCTGGCGAA 2443
QY 1343 CCCCGCGGGTTGAATTCGGGCTTCATGGGCGCACAGGTGACGGCGACCGGCTCCTGGC 1402
Db 2444 GAACTCGGGTTGAATTCGGGCTTCATGATCGCGAGGTGACCGCGCGGCTGGTGGC 2503
QY 1403 CGAGATGCGAGCAGCGGA---CCTGCTCGATTCATTCGATCTCCAGCAACCGCGCCAA 1459
Db 2504 CGAGTCCCGCTGCTGAGCCACCGCGTCCGTGATTCGATTCGCTCATCCGCGGCGCG 2563
QY 1460 TCAGGATGTGCTCGCTGGGACCATCGCGCGGCTCTGCGCGGAGAGATCGACCG 1519
Db 2564 AGAGGACACGTGTCCATGGGCGATGACGGCGGCTCAAGGGCGCTGAGTTCAGCGACT 2623
QY 1520 TTGGGCGGAGATCTTGGCATCTCGCTCTCTCTTGCACAAAGCTCGGAGCTGGCGCTG 1579
Db 2624 CGCCGCTGCTCGCTCGGATTAATCTGTGGCGGCGGCGGCTGAGCTTCGCGCT 2683
QY 1580 CGCAGCGGCTAGACGGGTGTCTCCGCGGGAAGAAGCTGGTCAGGCGCTCGCGGA 1639
Db 2684 GCGCTGAAGCCCGGAAGGCGCCCTCGCGGCTACGAGCTG-----TGCGCTC 2734
QY 1640 GCAGTTCCGCGCTTGAGAGCAGCGCGCCCTCGGAGCAGGAATTCGCGCTTGTCTAC 1699
Db 2735 GAAGGTCCTCCCATGATGACAAAGACCGGAGCTGACCGGAGCATCGAGCGGTGAGCCA 2794
QY 1700 GCACCTC 1706
Db 2795 GCTCGTC 2801
```

RESULT 12

ADS50013
ID ADS50013 standard; cDNA; 1503 BP.

XX
XX
AC ADS50013;

XX
XX
DT 02-DEC-2004 (first entry)

XX
XX
DE Bacterial polynucleotide #4756.

XX
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
pathogen tolerance; pest tolerance; plant disease resistance;
cell cycle pathway modification; plant growth regulator;
homologous recombination; seed oil yield; protein yield; carbohydrate;
nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
bacterial polynucleotide; gene; ss.

XX
OS Bacteria.

XX
PN US2003233675-A1.

XX
XX
PD 18-DEC-2003.

XX
XX
PF 20-FEB-2003; 2003US-00369493.

XX
XX

```
PR 21-FEB-2002; 2002US-0360039P.
XX (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
DR
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 28443; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 1503 BP; 261 A; 514 C; 476 G; 252 T; 0 U; 0 Other;
```

```
Query Match 11.4%; Score 211.6; DB 13; Length 1503;
Best Local Similarity 51.3%; Pred. No. 2e-31;
Matches 657; Conservative 0; Mismatches 569; Indels 54; Gaps 5;

QY 301 TCGGGCGCTGTCATCCGCGAGCGCGGCCATGTCTACGGACTGACAAACGGCTTCGGTCCCC 360
Db 125 TCGCCGATATCCCGGAAGGGGAGCGGCTTACGGCATCAACAGGGCTTTGGGGCC 184
QY 361 TTGCGAACCCCTGATCTCAGGTGAGAATGTCCGAACGCTGCAGGCCAATCTTGTCCATC 420
Db 185 TCGCCAGCAGCATATTCGCGACGACCAACTCGAACTGTCAGCGCAATCTGCTGCTCT 244
QY 421 ATCTGGCAGCGCGCTGGGACCGGTGCTTGACTGGAGCAGCGCGCGGCATGTTCTGG 480
Db 245 CTCACCGGTTCGGCGGTGGGTGAGCCGATGTCGGTCTCGGTCTGCTGTCATCGCGC 304
QY 481 CGCGTCTGCTGTCGATCTCAGGAGCGCTCGGTCGCCAGCGAGGGGACCATCGCTCGCC 540
Db 305 TGAACCTCTGAGCCTCGGCGCGCATTCGGGCATTTCGCGTGAAGTATGACGCGC 364
QY 541 TGATCGACCTGCTCAATTCGAGCTCGCTCCGCCCGTTCACAGCCCGCGACCGTGGCG 600
Db 365 TGATCAGCTGTACAAACCGGAGCTGCTCGCGGTGATTCCGGTCAAGGGTTTCGTCGGTG 424
QY 601 CGTGGGTGACCTGACACCGCTTGCGCATATGCTGCTCTGCCCTCCAGGGCGGGAGACT 660
Db 425 CATCGGGCGACCTCGCGCGCTCGCGCATATGTCGCGCGGCTGCTCGCGCTCGCGCA-- 482
```


Dbb 185 TCGCCAGCAGCATATTTCCGCA CGA CAACTCGAACTGTTGACGCGCAATCTGGTGTCT 244
QY 421 ATCTGGCCAGCGCGTGGGACCGGTGCTTGACTGAGACGACGCGCGCGCATGGTCTCTGG 480
Dbb 245 CTCA CGCGGTGCGGTGGGTGAGCCGATGTCGCGTCCGGTCTGCTCTGATCGCGC 304
QY 481 CGCGTCTGGTGTGATCTGCTCAGGAGCCTCCCGTCCACGAGGAGGACCATCGCTCGCC 540
Dbb 305 TGAATCTCTCGAGCTCGCGCGCGCCATTTCGGGCATTTCGCCGTGAAGTATGATGACGCGC 364
QY 541 TGATCGACCTGCTCAATTCGAGCTCGCTCGGCGGTTCCAGCGCGCGGACCGTGGCG 600
Dbb 365 TGATCA CGCTGTATCAACGCGGACGCTGCTCGCGGTGATTCGCGTCAAGGGTTCCGTCGGTG 424
QY 601 CGTCGGGTGACCTGACACGCGCTTTCGCGCATATGCTCTGCTCCAGGGCGCGGAGACT 660
Dbb 425 CATCGGCGACCTTCGCGCCGCTCGCGCATATGTCGGCGGCGCTGCTCGGGTTCGGCGA -- 482
QY 661 TCTTGAACCGGACGCGGACGCGCTTTGACGCGCGCAGAGGCTCCGCGCGGACGCGCTGC 720
Dbb 483 -AGTGTTCGCGAAGGCGAGCGCATGCGCGCCACCGAGGGTCTCGCGCTGTCGCGCTCA 541
QY 721 AACCGCTGATCTCTCCATCGCGATGCACTGGCGCTGTGTCACGCGGACCTCCGCGCATGA 780
Dbb 542 AGCCGCTCACGCTGCAAGCCAAAGAGGGGCTGGCGCTCTCTGAACGCGCAGCAGGCTTCGA 601
QY 781 CGGGATCGCGCTGGTGAATGCTCAGCGCTGCGCGCATCTCGGCAACTGGGCGGTGCGCT 840
Dbb 602 CGCGCTCGCGCTGTACAATGTTTCGCGCATCGAAGACCTGTACCGACCGCGCTGGTGT 661
QY 841 TGA CGGCGCTGCTTGGGGAATGCTGAGAGCGCGGACCGAGGCATGGCGCGCGCACTGT 900
Dbb 662 CGGCGCATTTGTCGGTAGATGCGCGAATGGCTCGGTCAAGCCGTTTCGACGCGCGCATTC 721
QY 901 CGACCTCGCGCGCATTCGCGGACAGAGAGACCGCGCAGGAGCTGCGGCGCGCGGTGG 960
Dbb 722 ACGAGTTGCGCGGCTCAAGGTGATGATCGACGCGCGCGGATACCGCTCGCTCTGCTGG 781
QY 961 ACGGACGCGCGGCTGCTCCGCGCTCATTTCCGCGCGGAGGCTCGACGCGCGGATA 1020
Dbb 782 AAGTTTCGCGCAATCAACGTTTTCGACG----- 808
QY 1021 TCGGGACGAGCGGAGCGGGGACGATGCTACAGCCCTGCGCTGCGCTCCGACGTTTC 1080
Dbb 809 -----CCGATTTCGACAGAGTGCAGGACCCGTACAGCTGCGCTGCCAGCCGCAAGTCA 862
QY 1081 TCGGGCGGCTTCGACACGCTCGCATGCGATGACCGGCTGCTGACGATGAGCTGAACG 1140
Dbb 863 TGGGCGGTGCTCTGGATCAGATCGCCACGCGGCCAACGTTGCTGCTCGAAGCGAACG 922
QY 1141 CGGTGACCGACATCCGCTGTTTCGCGCGATGGCAGCGTGCCCGCTGCGACGGGGGA 1200
Dbb 923 CGGTCTCCGACAACTCGCTGAATTTCCCGCA-----CACCAGCGAAGTGTGTGCGGCGGTA 979
QY 1201 ATTTCATGGCCAGCATGTGGCGCTGACGCTCGCATGCTCGCCACGCGCGTCAACGTTTC 1260
Dbb 980 ACTTCCACGAGACCGGTGCGCTTCGCGGCCGATTAACCTTGGCGTTCGCGCGCGCGAA 1039
QY 1261 TGGCGGCTTCGCGAGCGCGCATGTCAGCTCTGACAGTGAAGAGCTGAACCGTGGCG 1320
Dbb 1040 TCGCGCGCTTGGCGCAACGCGCATCGCGCTGCTGATCGACGCGAGCGTGTGCGGCGCTGC 1099
QY 1321 TGGCCCCCTTCCTCCACGGGGGCGCGCGGTGTAATTCGCGCTTCATGGGCGCACAG 1380
Dbb 1100 CGCC-----GTTTCCTGTGCGGATGGCGGCGTGAATTCGGGCTTCATGATCGCGCAG 1153
QY 1381 TGACGCGCACCGGCT---CCTGGCGAGATGGAGCCACGGACCTGCTCGTCCATCCATT 1437
Dbb 1154 TCAGGCTGGGCGCTCGCTCGGAAACAGACGCTCGCGCATCCGGCTTCGGTTCGATT 1213
QY 1438 CGATCTCCAGAACGCGCGCAATCAGGATGTGCTGCTGGGACCATCGCGCGCGCC 1497
Dbb 1214 CGTGTGCCACTTCGGCGAAACAGGAAGACCAACGTTGATGGGACAGTTTCGCGCGCGCA 1273

QY 1498 TCTGCGCGGAGAGATCGACCGTTGGCGGAGATCCTTGGATCCTCGTCTCTGCTTTG 1557
Dbb 1274 AGCTCGCGACATTGCTGAAACACCGCAACATTCTGTCGATCGAACTGCTCGCGCGC 1333
QY 1558 CACAAGCTCGGAGCTGCGC 1577
Dbb 1334 CGCAAGGCTCGATCTGCGC 1353

RESULT 14
ACA25783
ID ACA25783 standard; DNA; 1521 BP.
XX ACA25783;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #7440.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Burkholderia fungorum.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU21913.
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 13653; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX on a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for

Db	920	CGGTCTCCGACAAATCCGTGATTTTCCCCGA---CACCGGCGAAGTGCCTCGGGCGGTA	976
Qy	1201	ATTTCATGGGCCAGCATGTGGCGCTGACGTCCGATCGCTCCGACGGCCGTACACGTTTC	1260
Db	977	ACTTCCACGACAGCCGGTTCGGCGCGGATACCTCGCGTCTCGCCGCCCGCGGAAA	1036
Qy	1261	TGCGGGGCCCTTGCGGAGCGCCAGATTGCACGTCTGACAGATGAAGGCTGAACCGTGGGC	1320

DB TCGGCGCGCTGGCCGAAACGCCGCATCGCGCTGCTGATCGACGCCAGCGCTGTCGGGCTGC T096

[illegible][illegible]

Qy	CGATCTCCACGAAGCGCGCCAAATCAGATGTGGTCTCGCTTGGGACCATTCGCGCCGCGCC	1438
	1211	CGCTGCCCACTTCGGCAAAACGAGGAACCA
Db	CGCTGCGCACTTCGGCAAAACGAGGAACCA	1211
	CGATGTGATGGCGACGTTTCGCGCGCGCA	1270

1555
 1554
 1553
 1552
 1551
 1550
 1549
 1548
 1547
 1546
 1545
 1544
 1543
 1542
 1541
 1540
 1539
 1538
 1537
 1536
 1535
 1534
 1533
 1532
 1531
 1530
 1529
 1528
 1527
 1526
 1525
 1524
 1523
 1522
 1521
 1520
 1519
 1518
 1517
 1516
 1515
 1514
 1513
 1512
 1511
 1510
 1509
 1508
 1507
 1506
 1505
 1504
 1503
 1502
 1501
 1500
 1499
 1498
 1497
 1496
 1495
 1494
 1493
 1492
 1491
 1490
 1489
 1488
 1487
 1486
 1485
 1484
 1483
 1482
 1481
 1480
 1479
 1478
 1477
 1476
 1475
 1474
 1473
 1472
 1471
 1470
 1469
 1468
 1467
 1466
 1465
 1464
 1463
 1462
 1461
 1460
 1459
 1458
 1457
 1456
 1455
 1454
 1453
 1452
 1451
 1450
 1449
 1448
 1447
 1446
 1445
 1444
 1443
 1442
 1441
 1440
 1439
 1438
 1437
 1436
 1435
 1434
 1433
 1432
 1431
 1430
 1429
 1428
 1427
 1426
 1425
 1424
 1423
 1422
 1421
 1420
 1419
 1418
 1417
 1416
 1415
 1414
 1413
 1412
 1411
 1410
 1409
 1408
 1407
 1406
 1405
 1404
 1403
 1402
 1401
 1400
 1399
 1398
 1397
 1396
 1395
 1394
 1393
 1392
 1391
 1390
 1389
 1388
 1387
 1386
 1385
 1384
 1383
 1382
 1381
 1380
 1379
 1378
 1377
 1376
 1375
 1374
 1373
 1372
 1371
 1370
 1369
 1368
 1367
 1366
 1365
 1364
 1363
 1362
 1361
 1360
 1359
 1358
 1357
 1356
 1355
 1354
 1353
 1352
 1351
 1350
 1349
 1348
 1347
 1346
 1345
 1344
 1343
 1342
 1341
 1340
 1339
 1338
 1337
 1336
 1335
 1334
 1333
 1332
 1331
 1330
 1329
 1328
 1327
 1326
 1325
 1324
 1323
 1322
 1321
 1320
 1319
 1318
 1317
 1316
 1315
 1314
 1313
 1312
 1311
 1310
 1309
 1308
 1307
 1306
 1305
 1304
 1303
 1302
 1301
 1300
 1299
 1298
 1297
 1296
 1295
 1294
 1293
 1292
 1291
 1290
 1289
 1288
 1287
 1286
 1285
 1284
 1283
 1282
 1281
 1280
 1279
 1278
 1277
 1276
 1275
 1274
 1273
 1272
 1271
 1270
 1269
 1268
 1267
 1266
 1265
 1264
 1263
 1262
 1261
 1260
 1259
 1258
 1257
 1256
 1255
 1254
 1253
 1252
 1251
 1250
 1249
 1248
 1247
 1246
 1245
 1244
 1243
 1242
 1241
 1240
 1239
 1238
 1237
 1236
 1235
 1234
 1233
 1232
 1231
 1230
 1229
 1228
 1227
 1226
 1225
 1224
 1223
 1222
 1221
 1220
 1219
 1218
 1217
 1216
 1215
 1214
 1213
 1212
 1211
 1210
 1209
 1208
 1207
 1206
 1205
 1204
 1203
 1202
 1201
 1200
 1199
 1198
 1197
 1196
 1195
 1194
 1193
 1192
 1191
 1190
 1189
 1188
 1187
 1186
 1185
 1184
 1183
 1182
 1181
 1180
 1179
 1178
 1177
 1176
 1175
 1174
 1173
 1172
 1171
 1170
 1169
 1168
 1167
 1166
 1165
 1164
 1163
 1162
 1161
 1160
 1159
 1158
 1157
 1156
 1155
 1154
 1153
 1152
 1151
 1150
 1149
 1148
 1147
 1146
 1145
 1144
 1143
 1142
 1141
 1140
 1139
 1138
 1137
 1136
 1135
 1134
 1133
 1132
 1131
 1130
 1129
 1128
 1127
 1126
 1125
 1124
 1123
 1122
 1121
 1120
 1119
 1118
 1117
 1116
 1115
 1114
 1113
 1112
 1111
 1110
 1109
 1108
 1107
 1106
 1105
 1104
 1103
 1102
 1101

[illegible]

```

RESULT 15
AD50000
ID AD500000 standard; CDNA; 1518 BP.

```

XX
AC
XX
DT
AD50000;
02-DEC-2004
(first entry)

XX Bacterial polynucleotide #4743.
DE
XX
KW Recombinant DNA construct; transformed plant; improved plant property;

KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cold tolerance; heat tolerance; drought tolerance; nematode;
KW osmosis;
KW cell cycle pathway modification; plant growth regulator;
KW cell cycle pathway modification; seed oil yield; protein yield;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW

nitrogen; photosynthesis; lignin; galactomannan;
bacterial polynucleotide; gene; ss.
XX
OS Bacteria.

XX
PN US2003233675-A1.
XX
XX
PD 18-DEC-2003.

20-FEB-2003; 2003US-00369493.
21-FEB-2002; 2002US-0360039P.

XX
PA (CAOY//) CAO Y.
PA (HINK//) HINKLE G J.
PA (SLAT//) SLATER S C.

PA	CHEN X.
PA	(CHEN//)
PA	(GOLD//)
XX	GOLDMAN B S.
XX	
PI	Cao Y, Hinkle GJ, Slater SC, Goldman BS;

XX
XX
DR
XX
XX
PT
WPI; 2004-061375/06.
New recombinant DNA construct comprising a promoter positioned to provide

PT for expression of a polynucleotide encoding a polypeptide from a PT microbial source, useful for producing plants with improved properties.

XX
PS
XX

Claim 1; SEQ ID NO 28430; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polynucleotide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX
SQ
Sequence 1518 BP; 254 A; 504 G; 504 C; 256 T; 0 U; 0 Other;

Query Match 9.6%; Score 177.4; DB 13; Length 1518;

Best Local Similarity 49.5%; Pred. No. 7.7e-25;

Matches 688; Conservative 0; Mismatches 631; Indels 72; Gaps 6;

330 GTCTACGACTGACACCGGCTTCGGTCCCTTCGGAAACCGCCTGATCTCAGGTGAGAAAT 389
154 GTGTACGGCGTCAACACCGGTTTACGGCGATGCTCGCTGCTGATCGACGTCGATGGAAC 213
390 GTCCGAACCGTCAGAGCCCAATCTTGTCCATCATCTCGCCAGCGCGCTGGAGACCGGTGCTT 449
214 GTCCAGACACTGCGGTGACACTACCGGCTATACGGCTGCGGCATGAGGCCATGATCTC 273
450 GACTGGACGACGCGCGCGCATGGTCTTGGCGCGTCTGGTGTGATCGCTCAGGGAGCC 509
274 GACGAAGCGCAACCGTTCGGGTGATCGCGCGCGCTCAACTCGCTCGCATACGGTTT 333
510 TCGCGTCCAGCAGGGGACCATCGCTCGCTGATCGACCTGCTCAATTCGAGCTCGCT 569
334 TCGGGCGTGGCTCGGTGTGCTCGAAGCTCTCGCCGATCTGGTCAATCATCGCGTGTG 393
570 CCGGCGGTTCCAGCGCGACGCGTGGCGGTGCGGTGACCTGACACCGCTTGGCAT 629
394 CCGCGATTCGTCGGAAGGCTCGGTGCGCGGAGCGCGATCTCAGCGCGCTTTCGTAT 453
630 ATGGTGTCTCTCGCTCCAGGCGCGGGAGACTTCTCTGACCGGGACCGGACCGCGCTTGAC 689
454 GTGGCGCGCGCTCGCCGCTGAGCGGAGCTAATGTTGGAAGGGAGTTGCGCACTG 513
690 GGGCGAAGAGGGCTCCGGCGCGGACCGGTGCAACCGCTGATCTCTCCATCCGCGATGCA 749
514 CCGGAGGTGTGAGACCGAACTCGGTCAAAACG--CCGCTCACGCTCGCGCGCGAAGAGG 570
750 CTGGCGCTGTTCAACGGGACCTCCGCCATGACCGGGATCGCGTGGTGAATGCTCAGGCC 809
571 CTTCGCGCTGATGAACCGGACCGCGCTGATGACGGGTCTCGCCTGTCTCGGTTTGGCGC 630
810 TGCGCGCATCTCGGCAACTGGGCGGTGGGTTTGAACGCGCCCTGTCTCGGGAATGCTCTGAGA 869
631 GCCGATCATCTACGGGCTTACCGCGCGCTGACCGCGCTGTGCAAGGTTCGGCTCGAC 690
870 GGGCGGACCGAGGATGGGCGCGGCACTGTGCGAAGCTTCGCGCGCGCATCCCGGACAGAAG 929
691 GGCCTGTGCGCGCATTTTCGACGCAATGCTGTTTCGAGGTGAAGCCGACGCGCGCGCGCC 750

QY 930 GACGCCGACGCGAGGCTGCGCGCCGCTGTGGACGGCGAGCGCGCGGTGTGTCGGCAGTC 989
Db 751 GAAGCGCGCATGATCCGACGACGACCTGGCGGCGCGA----- 791
QY 990 ATTGCCGAGGGAGGCTCGACCGCGCGGATATCGGGACGGAGCGGAGGGGGGAGGAT 1049
Db 792 -----CGACACGCGCGGCGCACCGCTTTTTCAGGAC 819
QY 1050 GCCTACAGCTGCGCTGCGCTCCGAGGTTCTCGGGCGGGCTTTCGACACGCTCGCATGG 1109
Db 820 CGTATTTCGATTGCTGCGCGCGCGACGATCGGCGTGGCGCGCGACCGCTTTCGTGG 879
QY 1110 CATGACCGGCTGCTGACGATCGAGCTGAACGGGCTGACCGCAATCCGGTGTTCGCGCCC 1169
Db 880 GTACGCGCGACATCGAAGAACGAACTGAACAGCGCGAAGCAACCCGCTGATCGATCCG 939
QY 1170 GATGGAGGCTGCCCGCTTCGACCGGGGCAATTTTCATGGCGCAGATGTGGCGCTGACG 1229
Db 940 GA---CAACGAACGGCTGTGACCGCGCGCACTTTCAGCGCGCTCATATCGCTTTCGCG 996
QY 1230 TCCGATGCGCTGCCACGCGCTCACCGTTCTGGCGGGCTTTCGCGAGCGCCAGATTGCA 1289
Db 997 ATGGAATCGCTGAAGTCCGGTTCGCAACCTCGCGGATCTGATGACCGGCAACTGGCG 1056
QY 1290 CGTCTGACAGATGAAGGCTGAACCGTGGGCTGCCCGCTTCTCCACCG-----G 1340
Db 1057 CTCTCGTACGCTCAATTTCAATAACGGCTTGGCGCGCAATCTGTCCGGCGCTACGTCC 1116
QY 1341 GGGCCCGCGGGTTGAATTCGGCTTTCATGGCGGCAAGGTGAACGGCGACCGCGCTCCTG 1400
Db 1117 GCGCGCGCGCATCAATCAACGGGTTCAAGGCGGTGACAGATTCTCTCATCCGATGAGAC 1176
QY 1401 GCGCA---GATGGAGCAGCGGACCTGCTCGATCCATTGATCTCCAGGACGCGGCC 1457
Db 1177 GCGGAAGCATTTGAAGAACCAACATGCGCGCGAGGTTTCTCGCGCTCCACCGAGCGCAC 1236
QY 1458 AATCAGATGTGCTCGCTTGGGACCATCGCGCGCGCTCTGCGCGCAGAGAGATCGAC 1517
Db 1237 ACCAGGACAAAGTCAGCATGGCAGCATCGCGCGCGGAGTGCCTGCGCTGCTGGA 1296
QY 1518 CGTGGCGGAGATCCTTCGATCTCTGCTCTGTTGCAAGCTCGGAGCTGCGC 1577
Db 1297 CTGACCGAACAGGTGGCGGCGAGCGCATACGCTCGCAACCGTGCAGCTGCCGCTCTGCGA 1356
QY 1578 TGGCGCAGCG---GCTAGACGGGTGTCTCCCGGGGGAAGAACTGTCGAGGCGCCTG 1634
Db 1357 TTGAAGATCGATAGCGGAACACCGGTTCCCGCGCGCTCGGAACGTTCTATGAGAGCGTG 1416
QY 1635 CCGGAGCAGTTCCCGCGCTTGAAGACCGGACCGCGCCCTCGGACAGGAAATTCGCGCGCTT 1694
Db 1417 AGCGCGAGTTCGCGGTTCTGAGACGAGACCGCGCGCTCGAACACGAACTGCGCGCCTG 1476
QY 1695 GTACGCACT 1705
Db 1477 ACGCGCGCAT 1487

Search completed: December 10, 2005, 12:35:04
Job time : 1140 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:06:26 ; Search time 186 Seconds
(without alignments)
1242.544 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLANSPPKPAVELDRHI.....RPLQGEIATLALHLLQQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A: Geneseq_21.*

1: geneseqp1980s.*

2: geneseqp1990s.*

3: geneseqp2000s.*

4: geneseqp2001s.*

5: geneseqp2002s.*

6: geneseqp2003as.*

7: geneseqp2003bs.*

8: geneseqp2004s.*

9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	100.0	526	8 ADJ57136	Adj57136 R. sphaer
2	2444	91.6	482	8 ADN25163	Adn25163 Bacterial
3	1335	50.0	539	8 ADL16888	Adl16888 Rhodobact
4	1335	50.0	540	8 ADL16877	Adl16877 Rhodobact
5	1333	50.0	540	8 ADL16887	Adl16887 Rhodobact
6	837	31.4	529	8 ADN26159	Adn26159 Bacterial
7	814.5	30.5	539	3 AAB13586	Aab13586 Streptomy
8	761	28.5	508	9 ABM95993	Abm95993 M. xanthu
9	755	28.3	501	8 ADN22103	Adn22103 Bacterial
10	755	28.3	501	8 ADN24862	Adn24862 Bacterial
11	755	28.3	507	6 ASU21913	Asu21913 Protein e
12	753.5	28.2	501	8 ADS30178	AdS30178 Bacterial
13	749.5	28.1	507	6 ASU22102	Asu22102 Protein e
14	748	28.0	522	8 ADN26467	Adn26467 Bacterial
15	734.5	27.5	502	8 ADN24391	Adn24391 Bacterial
16	734.5	27.5	511	6 ASU21162	Asu21162 Protein e
17	729.5	27.3	487	8 ADN21635	Adn21635 Bacterial
18	729	27.3	460	8 ADS21195	AdS21195 Bacterial
19	728.5	27.3	508	8 ADS44935	AdS44935 Bacterial
20	715	26.8	494	8 ADS24598	AdS24598 Bacterial
21	712	26.7	511	8 ADS28241	AdS28241 Bacterial
22	711.5	26.7	524	8 ADS45022	AdS45022 Bacterial
23	709	26.6	506	8 ADN24850	Adn24850 Bacterial
24	709	26.6	506	8 ADN22090	Adn22090 Bacterial

25	708.5	26.6	511	6 ABU49321	Abu49321 Protein e
26	706	26.5	496	8 ADS23085	AdS23085 Bacterial
27	703.5	26.4	485	8 ADS21886	AdS21886 Bacterial
28	699.5	26.2	484	8 ADN25988	Adn25988 Bacterial
29	695.5	26.1	657	8 ADP24794	Adp24794 PRO polyp
30	695	26.0	510	8 ADS22467	AdS22467 Bacterial
31	694.5	26.0	657	8 ADN04255	Adn04255 Antipbort
32	694.5	26.0	657	8 ADP23988	Adp23988 PRO polyp
33	694	26.0	497	8 ADN24930	Adn24930 Bacterial
34	688.5	25.8	511	6 ABU16504	Abu16504 Protein e
35	688	25.8	507	8 ADS26381	AdS26381 Bacterial
36	688	25.8	507	8 ADS27132	AdS27132 Bacterial
37	688	25.8	507	8 ADS26749	AdS26749 Bacterial
38	687	25.7	482	8 ADN22171	Adn22171 Bacterial
39	687	25.7	525	8 ADL05169	AdL05169 M. catar
40	683.5	25.6	491	4 AAU34175	Aau34175 Staphyloc
41	683.5	25.6	504	4 AAU37099	Aau37099 Staphyloc
42	683.5	25.6	556	4 AAU43358	Aau43358 Propionib
43	683.5	25.6	556	6 ABM39877	Abm39877 Propionib
44	683	25.6	508	7 ABO67058	AbO67058 Klebsiell
45	683	25.6	519	6 ABU35328	Abu35328 Protein e

ALIGNMENTS

RESULT 1

ID	ADJ57136	ADJ57136 standard; protein, 526 AA.
XX	AC	ADJ57136;
XX	DT	06-MAY-2004 (first entry)
XX	DE	R. sphaeroides tyrosine ammonia lyase (TAL) enzyme.
XX	KW	TAL; tyrosine ammonia lyase; catalyst; para-hydroxycinnamic acid; PHCA;
XX	KW	tyrosine; enzyme.
XX	OS	Rhodobacter sphaeroides.
XX	PN	WO2004009795-A2.
XX	PD	29-JAN-2004.
XX	PF	23-JUL-2003; 2003WO-US023229.
XX	PR	23-JUL-2002; 2002US-0397820P.
XX	FA	(DUPO) DU PONT DE NEMOURS & CO E I.
XX	PI	Huang L, Xue Z;
XX	DR	WPI; 2004-214331/20.
XX	DR	N-PSDB; ADJ57135.
XX	PT	Novel tyrosine ammonia lyase enzyme, used as catalyst in producing para-hydroxycinnamic acid from tyrosine.
XX	PS	Claim 5; SEQ ID NO 3; 53pp; English.
XX	CC	The invention relates to a tyrosine ammonia lyase enzyme (TAL) and encoding polynucleotides. The TAL polypeptide is used as a catalyst in producing para-hydroxycinnamic acid (PHCA) from tyrosine. The method involves contacting a transformed host cell expressing the polypeptide with a fermentable carbon substrate, e.g. monosaccharides, oligosaccharides, polysaccharides, carbon dioxide, methanol, formaldehyde, formate and carbon-containing amines, preferably glucose, growing the host cell for a time sufficient to produce PHCA, and optionally recovering PHCA. The TAL polypeptide enables simple and cost-effective production of PHCA. The present sequence represents a R. sphaeroides TAL polypeptide.

SQ Sequence 526 AA;

Query Match 100.0%; Score 2668; DB 8; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.5e-220;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKPWLMSPPKPAVELDRHIDLDQAHAVASGARIVLAPPARDRCRASEARLGVAREAR 60
DB 1 VKPWLMSPPKPAVELDRHIDLDQAHAVASGARIVLAPPARDRCRASEARLGVAREAR 60
QY 61 HVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLDWTTRAMVRLARLVSAQG 120
DB 61 HVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLDWTTRAMVRLARLVSAQG 120
QY 121 ASGASEGTIARLIDLINSELAPAVPSRGTGASGDLTFLAHMVLCLQGRGDFLDRDGTGL 180
DB 121 ASGASEGTIARLIDLINSELAPAVPSRGTGASGDLTFLAHMVLCLQGRGDFLDRDGTGL 180
QY 181 DGAELRRGRLOPLDLSHRDALALVNGTSAMTGIALVNAHACRHGNWAVALTALLAECL 240
DB 181 DGAELRRGRLOPLDLSHRDALALVNGTSAMTGIALVNAHACRHGNWAVALTALLAECL 240
QY 241 RGRTEAWAALSILRPHGQKDAARLARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ 300
DB 241 RGRTEAWAALSILRPHGQKDAARLARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ 300
QY 301 DAYSLRCAPOVLGAGFDTLAWHDRVLTIELNAVTDNVPFPGDGVPAHLHGNFMGQHV 360
DB 301 DAYSLRCAPOVLGAGFDTLAWHDRVLTIELNAVTDNVPFPGDGVPAHLHGNFMGQHV 360
QY 361 TSDALATAVTVLAGLAEQRIARLDERLNRLGPPFLHGRGAGLNSGFMGQVATALLAE 420
DB 361 TSDALATAVTVLAGLAEQRIARLDERLNRLGPPFLHGRGAGLNSGFMGQVATALLAE 420
QY 421 MRATGPASIHSTNAANQDVVSGTIAARLCREKIDRWAEILAILALCLAQAEELRCGS 480
DB 421 MRATGPASIHSTNAANQDVVSGTIAARLCREKIDRWAEILAILALCLAQAEELRCGS 480
QY 481 GLDGVSPAGKKLVQALREQFPPLTDRPLGQEIATLHLLQQSPV 526
DB 481 GLDGVSPAGKKLVQALREQFPPLTDRPLGQEIATLHLLQQSPV 526

RESULT 2
ADN25163
ID ADN25163 standard; protein; 482 AA.
XX ADN25163;
XX ADN25163;

XX 02-DEC-2004 (first entry)
XX Bacterial polypeptide #7816.
XX Recombinant DNA construct; transformed plant; improved plant property;
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX pathogen tolerance; pest tolerance; plant disease resistance;
XX cell cycle pathway modification; plant growth regulator;
XX homologous recombination; seed oil yield; protein yield; carbohydrate;
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX bacterial polypeptide.
XX Bacteria.
XX OS
XX PN
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.

PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
DR New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 7816; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX Sequence 482 AA;
SQ Query Match 91.8%; Score 2444; DB 8; Length 482;
Best Local Similarity 100.0%; Pred. No. 1.2e-200;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 DRCRASEARLGVAREARHVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLD 102
DB 1 DRCRASEARLGVAREARHVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLD 60
QY 103 WTTARAVMLARLVSIAGGASGASEGTIARLIDLINSELAPAVPSRGTGASGDLTFLAHM 162
DB 61 WTTARAVMLARLVSIAGGASGASEGTIARLIDLINSELAPAVPSRGTGASGDLTFLAHM 120
QY 163 VLCLQGRGDFLDRDGTGLRDLGAEGLRRGRLOPLDLSHRDALALVNGTSAMTGIALVNAHAC 222
DB 121 VLCLQGRGDFLDRDGTGLRDLGAEGLRRGRLOPLDLSHRDALALVNGTSAMTGIALVNAHAC 180
QY 223 RHLGNAVALTALLAECLGRTEAWAALSILRPHGQKDAARLARVDGSRVVRHV 282
DB 181 RHLGNAVALTALLAECLGRTEAWAALSILRPHGQKDAARLARVDGSRVVRHV 240
QY 283 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFDTLAWHDRVLTIELNAVTDNVPFPPD 342
DB 241 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFDTLAWHDRVLTIELNAVTDNVPFPPD 300
QY 343 GSVPALHGNFMGQHVALTSDALATAVTVLAGLAEQRIARLDERLNRLGPPFLHGRGAG 402
DB 301 GSVPALHGNFMGQHVALTSDALATAVTVLAGLAEQRIARLDERLNRLGPPFLHGRGAG 360
QY 403 LNSGFMGAQVTTALLAEWRATGPASIHSTNAANQDVVSGTIAARLCREKIDRWAEI 462
DB 361 LNSGFMGAQVTTALLAEWRATGPASIHSTNAANQDVVSGTIAARLCREKIDRWAEI 420

QY 463 LAILALCLAAAEELRCGSLDGVSPAGKXIVQALREQFPFLETRDPLGQEIATLATHLIQ 522
DB 421 LAILALCLAAAEELRCGSLDGVSPAGKXIVQALREQFPFLETRDPLGQEIATLATHLIQ 480
QY 523 QS 524
DB 481 QS 482
RESULT 3
ADL16888
ID ADL16888 standard; protein; 539 AA.
AC ADL16888;
XX
DT 06-MAY-2004 (first entry)
DE Rhodobacter capsulatus TAL mutant protein.
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; enzyme;
KW mutant.
XX
OS Rhodobacter capsulatus.
XX Synthetic.
XX
PN US2004029230-A1.
XX
PD 12-FEB-2004.
XX
PF 18-JUN-2003; 2003US-00464609.
XX
PR 18-JUN-2002; 2002US-0389593P.
XX
PA (KYND/) KYNDT J J A.
PA (BEEU/) BEEUMEN J V.
XX
PI Kyndt JJA, Beeumen JV;
XX WPI; 2004-168893/16.
XX
XX New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
PT protein used as photochromic substance in electro optical random access
PT memory or in biochips.
XX
PS Example 1; Page; 32pp; English.
XX
CC The present invention is generally related to recombinant DNA technology
CC and particularly to DNA strands useful for the production of
CC parahydroxycinnamic acid and photoactive yellow protein (PYP) in a
CC suitable host expression system. The invention is useful for synthesizing
CC parahydroxycinnamic acid and derivative products such as holo-photoactive
CC yellow protein which is useful as a photochromic substance in electro
CC optical random access memory, in biochips, in light-inducible gene
CC expression systems and in liquid crystal polymers. The present sequence
CC is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) mutant protein.
CC The mutant protein is constructed with the deletion of Ala at position
CC 535. Note: This protein is shown in the specification but is derived from
CC Rhodobacter capsulatus wild type tyrosine ammonia lyase (TAL) protein
CC shown as SEQ ID NO 4 in page 13-14 of this specification.
XX
SQ Sequence 539 AA;
Query Match 50.0%; Score 1335; DB 8; Length 539;
Best Local Similarity 56.5%; Pred. No. 2.1e-105;
Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;
QY 11 KPAVELDRHIDIDQAHAVASGGARIVLAPPARDRCRASEARLGVIREARHYGLTTGFG 70
DB 19 KDCIALDGLATLVQCEATATHRISRTVTPALRRCARAHLEAIAEQRIHYGITGFG 78
QY 71 PLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARLVSIAGGASGSEGTTA 130

DB 79 PLANRLIADQGAELQQNLIIYHLATGVGPXKLSWAERALMLARLNSIIQGASGSPETID 138
QY 131 RLIDLLNSELAPVPSRGTVGASGDLTFLAHMVLCLQGRGDFLDRDGTGLDCAEGLRRGR 190
DB 139 RIVAVLNAGFAPEVPAQGTVGASGDLTFLAHMVLALQGRMIDPSGRVQEAQAVMDRLC 198
QY 191 LOPDLSHRDALALVNGTSMTGI-ALVNAHACRHLGNWVALTALLAECLRGRTEAMAA 249
DB 199 GGPLTAAARDGLALVNGTSMTATAALTGVEAARAI-DAALRHSAVLMEVLSGHAEAWHP 257
QY 250 ALSDLRPHPGOKDRAARLRARVDGSRVVRVIAERRLDAGDIGTEPEAGODAYSIRCAP 309
DB 258 AFAEURPHPGQLRATERLAQALDGAGRVCRITLTAARRLTAADLRPEHPAQDAYSLRVVP 317
QY 310 QVLGAGFDTLAWHDRVLTIELNAVTDNPFPPDGSVPALHGGNFNGQHVALTSDALATAV 369
DB 318 QLVGAVMTLDWHDVVVTCELNSVTDNPIFPEGCAVPALHGGNFNGVHVALASDALNAAL 377
QY 370 TVLAGLABRQIARLTDERLNKGLPPFLHRGPAGLNSGFMGAQVTTATALLAENRATG-PAS 428
DB 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGGQAGLQSGFMGAQVTTATALLAENRANATPVS 437
QY 429 IHSISTNAANODVSLGTIARLCREKIDRWAEILAILALCLAAAEELRCG-SGLDGVSP 487
DB 438 VQSLSTNGANQDVVSMGTIAARRAQLPLSQIAILALALAAQAWDLDDPEQGQWSL 497
QY 488 AGKKLVQALREQFPFLETRDPLGQEIATLATHLIQ 524
DB 498 TARDLRORIRAVSPGLRADRPLAGDIEAVALQRLRHS 534
RESULT 4
ADL16877
ID ADL16877 standard; protein; 540 AA.
XX
AC ADL16877;
XX
DT 06-MAY-2004 (first entry)
XX
DE Rhodobacter capsulatus tyrosine ammonia lyase (TAL) protein.
KW Para-hydroxycinnamic acid; photoactive yellow protein;
KW photochromic substance; biochip; liquid crystal polymer; PYP; enzyme.
XX
OS Rhodobacter capsulatus.
XX
FH Key Location/Qualifiers
FT Misc-difference 534..535 /note= "Encoded by TCNNNGCC"
XX
US2004029230-A1.
XX 12-FEB-2004.
XX 18-JUN-2003; 2003US-00464609.
XX 18-JUN-2002; 2002US-0389593P.
XX (KYND/) KYNDT J J A.
XX (BEEU/) BEEUMEN J V.
XX Kyndt JJA, Beeumen JV;
XX WPI; 2004-168893/16.
XX N-PSDB; ADL16876.
XX
XX New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
PT protein used as photochromic substance in electro optical random access
PT memory or in biochips.
XX
PS Claim 23; SEQ ID NO 4; 32pp; English.

XX The present invention is generally related to recombinant DNA technology
 CC and particularly to DNA strands useful for the production of
 CC parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
 CC suitable host expression system. The invention is useful for synthesizing
 CC parahydroxycinnamic acid and derivative products such as holo-photoactive
 CC yellow protein which is useful as a photochromic substance in electro
 CC optical random access memory, in biochips, in light-inducible gene
 CC expression systems and in liquid crystal polymers. The present sequence
 CC is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) protein.
 XX
 SQ Sequence 540 AA;

Query Match 50.0%; Score 1335; DB 8; Length 540;
 Best Local Similarity 56.5%; Pred. No. 2.1e-105;
 Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;

Qy 11 KPAVELDRHIDLDQAHAVAGGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFG 70
 Db 19 KCCLALDGLTLVQCEAIATHRSRISVTPALRERCARAHARLEHAIAEQRHIYGITGFG 78
 Qy 71 PLANRLISGENVRLTQANLVHILASGVGPVLDWTTARAVLRLVSTAOAGSASGETIA 130
 Db 79 PLANRLIGADQGAELQONLIYHLATGVPKLSWAERARLMLRLNSILOAGSPETID 138
 Qy 131 RLIDDLNSELAPAVPSRGTGASGDLTPLAHMVLCIQGRGDFLDRDGTDLDAEGLRRGR 190
 Db 139 RIVAVLNAGFAPEVPAQGTGASGDLTPLAHMVLCIQGRGDMIDPSGRVQAGVMDRLC 198
 Qy 191 LQPLDLSHRDALALVNGTSAMTGI-ALVNAHACRHNLGNVAVALTALLAECLRGTEAWAA 249
 Db 199 GGPLETLAARDGLALVNGTSAMTGI-ALVNAHACRHNLGNVAVALTALLAECLRGTEAWAA 257
 Qy 250 ALSDLRPHGQKDAARLARVDGSRVVRHVTAERRLDAGDGTGTEPEAGQDAYSRLCAP 309
 Db 258 AFALRPHGQKDAARLARVDGSRVVRHVTAERRLDAGDGTGTEPEAGQDAYSRLCAP 317
 Qy 429 IHSISTNAANQDVVSLGTTIARLCKRIDRWABILAILALCLAOAAELRCG-SGLDQVSP 487
 Db 438 VQSLSTNGANQDVVSLGTTIARLCKRIDRWABILAILALCLAOAAELRCG-SGLDQVSP 497
 Qy 488 AGKKLVQALREQFPFLETDRPLGQEIATLALATHLLQSS 524
 Db 498 TARDLRIRAVSGLRADPLAGDIEAVAGLRRHS 534

RESULT 5
 ADL16887
 ID ADL16887 standard; protein; 540 AA.
 XX
 AC ADL16887;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Rhodobacter capsulatus TAL mutant protein (D522H).
 XX
 KW Para-hydroxycinnamic acid; photoactive yellow protein;
 KW photochromic substance; biochip; liquid crystal polymer; pyp; enzyme;
 KW muten; mutant.
 XX
 OS Rhodobacter capsulatus.
 OS Synthetic.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 522

FT XX /note= "wild type Asp replaced with His"
 PN US2004029230-A1.
 XX
 PD 12-FEB-2004.
 XX
 PF 18-JUN-2003; 2003US-00464609.
 XX
 PR 18-JUN-2002; 2002US-0389593P.
 XX
 PA (KYND/) KYNDT J J A.
 PA (BEEU/) BEEUMEN J V.
 XX
 KYNDT JJA, Beeumen JV;
 XX
 WIPI; 2004-168893/16.
 DR
 XX New nucleic acid encoding tyrosine-ammonium lyase polypeptide, for
 PT synthesizing para-hydroxycinnamic acid and holo-photoactive yellow
 PT protein used as photochromic substance in electro optical random access
 PT memory or in biochips.
 XX
 PS Example 1; Page; 32pp; English.
 XX
 CC The present invention is generally related to recombinant DNA technology
 CC and particularly to DNA strands useful for the production of
 CC parahydroxycinnamic acid and photoactive yellow protein (pyp) in a
 CC suitable host expression system. The invention is useful for synthesizing
 CC parahydroxycinnamic acid and derivative products such as holo-photoactive
 CC yellow protein which is useful as a photochromic substance in electro
 CC optical random access memory, in biochips, in light-inducible gene
 CC expression systems and in liquid crystal polymers. The present sequence
 CC is Rhodobacter capsulatus tyrosine ammonia lyase (TAL) mutant protein.
 CC Note: This protein is shown in the specification but is derived from
 CC Rhodobacter capsulatus wild type tyrosine ammonia lyase (TAL) protein
 CC shown as SEQ ID NO 4 in page 13-14 of this specification.
 XX
 SQ Sequence 540 AA;

Query Match 50.0%; Score 1333; DB 8; Length 540;
 Best Local Similarity 56.5%; Pred. No. 3.2e-105;
 Matches 292; Conservative 58; Mismatches 163; Indels 4; Gaps 4;
 Qy 11 KPAVELDRHIDLDQAHAVAGGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFG 70
 Db 19 KCCLALDGLTLVQCEAIATHRSRISVTPALRERCARAHARLEHAIAEQRHIYGITGFG 78
 Qy 71 PLANRLISGENVRLTQANLVHILASGVGPVLDWTTARAVLRLVSTAOAGSASGETIA 130
 Db 79 PLANRLIGADQGAELQONLIYHLATGVPKLSWAERARLMLRLNSILOAGSPETID 138
 Qy 131 RLIDDLNSELAPAVPSRGTGASGDLTPLAHMVLCIQGRGDFLDRDGTDLDAEGLRRGR 190
 Db 139 RIVAVLNAGFAPEVPAQGTGASGDLTPLAHMVLCIQGRGDMIDPSGRVQAGVMDRLC 198
 Qy 191 LQPLDLSHRDALALVNGTSAMTGI-ALVNAHACRHNLGNVAVALTALLAECLRGTEAWAA 249
 Db 199 GGPLETLAARDGLALVNGTSAMTGI-ALVNAHACRHNLGNVAVALTALLAECLRGTEAWAA 257
 Qy 250 ALSDLRPHGQKDAARLARVDGSRVVRHVTAERRLDAGDGTGTEPEAGQDAYSRLCAP 309
 Db 258 AFALRPHGQKDAARLARVDGSRVVRHVTAERRLDAGDGTGTEPEAGQDAYSRLCAP 317
 Qy 310 OVLGAGFDTLAWHDRVLTIELNAVTONPVPPDGSVPALHGGNFMGOHVALTSDALATAV 369
 Db 318 QLVGAVWDTLDHDRVVTCELNSVTNPIPEGCAPALHGGNFMGOHVALTSDALATAV 377
 Qy 370 TVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PAS 428
 Db 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGGQAGLQSGFMGAQVTTATALLAEMRANATPVS 437
 Qy 429 IHSISTNAANQDVVSLGTTIARLCKRIDRWABILAILALCLAOAAELRCG-SGLDQVSP 487
 Db 438 VQSLSTNGANQDVVSLGTTIARLCKRIDRWABILAILALCLAOAAELRCG-SGLDQVSP 497

CC XX

CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 507 AA;

Query Match 28.3%; Score 755; DB 6; Length 507;
Best Local Similarity 38.8%; Pred. No. 1.2e-55;
Matches 199; Conservative 73; Mismatches 211; Indels 30; Gaps 11;

Qy 19 HIDLDQAHAVASGGARIVLAP---PARDRCRSEARLGAIVREARHVYGLTTGFGPLANR 75
Db 8 HLTLPLQLRIAREHVALQLDPASHAAIDACAQAVADIAA---KGEPAVGINTGFGRLAST 64
Qy 76 LISENVRTTQANLVHHLASGVPLDWTTRAMVRLARLVSIAGASASGTTARLIDL 135
Db 65 HIPDQLELLQRNLVLSHAVGVGPFMSRPVVRLLITALKSLSLGRGSHGIRREVMDALITL 124
Qy 136 LNSELAPAVPSRGTVGASGDLTPLAHMVLCLQSGRGLDRLDGTGLDGAEGLRRCRLOPLD 195
Db 125 YNADVLPVPIPVKGSVGASGDLAPLAHMSAALLGVGEVFAK--GERMPATEGALVGLKPLT 183
Qy 196 LSHRDALALVNGTSMATGIALVNAHAHRHLGNWAVALTALLAECLIRGRTEAWAALSCLR 255
Db 184 LQAKEGLALLNGTQASTALALYNFPAIEDLYRTALVSGALSVDAAMSGVKFPDARIHELRL 243
Qy 256 PHFGQKDAARLRLRVGSGARVRHVIAERRLDAGDIGTEPEAGODAYSRLRCAPQVLGAG 315
Db 244 GHQCQIDAAAYRSLLEGSAINVSHADCDK-----VQDPYSLRCQPQVMGAC 290
Qy 316 FDTLAWHRDLVLTIELNAVTPNP--VPPDGSYPALHGGNFMGQHVHVALTSDALATATVVLG 374
Db 291 LDQMRHAANVLLLEANAUSDNPLFPDPTGEV--LSGGNFHAEPVAFADNLALAAAEIGA 348
Qy 375 LAERQIARLTDERLNRLGPPPLHRLGPGAGLNSGFGMAOVTAALLAEAMRATG-PASIHISIS 433
Db 349 LAERIALLLIDATLIS-GLPPFLVR-DGVNSGFMIAHVTAALAASENKTLAHPASVDSLP 406
Qy 434 TNAANQDVSLGTTAARLCREKIDRWABEILAILALCLAQAAELRCGSLDGVSPAGKKLV 493
Db 407 TSANQEDHVSMTFAARKLGIDIAENTANILSIELLAAAGQVDLRAP--HKTSPSLQKAM 463
Qy 494 QALREQPPPLETDRPLQGEIAALATHLLQQSPV 526
Db 464 DTRVKDVAHYELDHYPADIAAV--TRLVQNGTI 495

RESULT 12
ADS30178
ID ADS30178 standard; protein; 501 AA.
XX AC
XX ADS30178;
XX
XX
DT 02-DEC-2004 (first entry)
XX
XX Bacterial polypeptide #19211.
DE
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
XX
XX US2003233675-A1.
XX PN

18-DEC-2003.

20-FEB-2003; 2003US-00369493.

21-FEB-2002; 2002US-0360039P.

(CAOY/) CAO Y.
(HINK/) HINKLE G J.
(SLAT/) SLATER S C.
(CHEN/) CHEN X.
(GOLD/) GOLDMAN B S.

Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
WPI; 2004-061375/06.

New recombinant DNA construct comprising a promoter positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source, useful for producing plants with improved properties.

Claim 1; SEQ ID NO 19211; 122pp; English.

The invention relates to a recombinant DNA construct comprising a promoter functional in a plant cell, where the promoter is positioned to provide for expression of a polynucleotide encoding a polypeptide from a microbial source. The invention also relates to a transformed plant comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 501 AA;

Query Match	28.2%	Score 753.5;	DB 8;	Length 501;
Best Local Similarity	39.0%	Pred. No. 1.5e-55;		
Matches 196: Conservative	81;	Mismatches 197;	Indels 29;	Gaps 9

Qy	20	IDLQAHAVASGGGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGPGLPNRLISG	79
Dd	8	LKLEIILQVARNEAETVELSPDAATRVASRALVDRAAGDTPAYGINTGFPGFLAEVRIDK	67
Qy	80	ENVRTTLQANLVHHLASGVGPVLDTWTARTAMVARLVSTAOCSGASECTIARLLDLLNSE	139
Dd	68	KDLRDQRNIIISHACGVGTPLPIPEARALLJLRNCNLAKSYSGIRMETFLALADMLNRD	127
Qy	140	LAPAVPSRGTVGSAGDUTPILAHMVLCHOGRGDFLDROGDGAELGRRLQPQLDLSHR	199
Dd	128	VVPVPPERGVSAGSDLAHLAHLALVFIGEAF-YQGORMPAKOALERAGLOPVVLEAK	186
Qy	200	DALAIVNGTSAMTIGIALVNHAACHRLGNWVALTALLAACLCURGERTEAAWAALSCLRPHG	259
Dd	187	EGLALVNGTQMCAVGITLLQLRAESLADIADVAGMTLEGLLGSHKPFIPEIHDRVRAHPG	246
Qy	260	QKDAAELRARVDGSGARVVRRHVIAERRLLDAGDIGTEPEGDAQYSRLCAPQVLGAGFTDL	319
Dd	247	QKDVAHLHRLIIVDSSELVESFVNSCK-----VDDPYSLRWCPQHVGAAREGI	293

RESULT 12
ADCS30178

AD530178
ID AD530178 standard; protein; 501 AA.

XX

AC ADS30178;

3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527

DT 02-DEC-2004 (first entry)
yyXX
DE
Bacterial polymeptide #19211

DE
XX
vastergat potuerepide #19211:

Recombinant DNA construct; transformed plant; improved plant property; cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic; pathogen tolerance; pest tolerance; plant disease resistance; call cycle pathway modification; plant growth regulator; homologous recombination; seed oil yield; protein yield; carbohydrate; nitrogen; phosphorus; photosynthesis; lignin; galactomannan; bacterial polypeptide.

13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 104

OS Bacteria.

[illegible]

PN yy

comprising the recombinant DNA construct and a method of producing a transformed plant having an improved property. The plant is a crop plant such as maize or soybean. The method of producing a transformed plant having an improved property comprises transforming a plant with the recombinant DNA construct and growing the transformed plant, where the polynucleotide or polypeptide is useful for improving plant properties. The recombinant DNA construct is useful for producing plants with improved plant properties, e.g. improved cold, heat or drought tolerance, tolerance to herbicides, extreme osmotic conditions, pathogens or pests, increased resistance to plant disease, better growth rate by modification of the cell cycle pathway with plant growth regulators, increased rate of homologous recombination, modified seed oil or protein yield and/or content, improved yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, by modification of photosynthesis or by providing improved plant growth and development under at least one stress condition, improved lignin production or improved galactomannan production. This sequence represents a bacterial polypeptide used in the scope of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

Sequence 502 AA;

Query Match		27.5%;	Score 734.5;	DB 8;	Length 502;
Best Local Similarity		38.3%;	Pred. No. 6.5e-54;		
Matches 193;		Conservative 73;	Mismatches 217;	Indels 21;	Gaps 7;
Qy	19	HIDLDOAHAVASGGARIVLAPPARDRCRSEARLGAVIREARHVYGLTTGFGPLANRLIS	78		
Db	10	HLTLAQRRIAHFSETVOLCFSTFAADASAQAVENIVRKGPVYGVNTGFRLASTHIP	69		
Qy	79	GENVRTLQANLVHHLASGVPLDWTTRAMVLARLVIAQAGSASGEGTARLIDLINS	138		
Db	70	LEQDLLQRLNVLVSHAVGVEPMSRVVRLMIVLKSSLSRGHSIRREVIALVTLNA	129		
Qy	139	ELAPAPVSRGTVGASGDLTPLAHNVLCLOGRGDFLDGRDTRLDGAEGLRRGRLOPLDLSH	198		
Db	130	DVLPVPIVKGSVGASGLAPLAHMSAVLLGVGEVTIGE-RRASANEGLHAAGLEPFTLRA	188		
Qy	199	RDALVNGTSMGTGIALVNAHACRHGNWAVALTALLAECLRGRTAWAAALSDLRPHP	258		
Db	189	KEGLALNGTQASTALAFNFAIEDLFTALVAGALSVDAAAASVVFPDARIHALRGHQ	248		
Qy	259	GOKDAARLRARVDGARVVRHVITAERRLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDT	318		
Db	249	GQDSAAAYRALLEGSINLSH-----SDCKVQDPYSURCOPQVWGACLHQ	295		
Qy	319	LAWHDRVLTIELNAVTDNPVFPDGSVPALHGGNFMGQHVALTSDALATATVTVLAGLAER	378		
Db	296	MRHSAEVLCEANSVSDNPLIFPD-TCDLVSGGNFHAEPVAFADNLAVALAEIGALAE	354		
Qy	379	QIARLTDERLNRLGPPFLHGRPGAGLNGFGMAQVATALLAEWRATG-PASIHISISTNAA	437		
Db	355	RIALLIDTTLIS-GLPPEFLVR-DSGLNSGFMIAHVTAALASENKTLAHPASVDSLPTSAN	412		
Qy	438	NQDVVSLGTIAARLCKEKIDRWAEILAILALCLAAAELECGSLDGVSPAGKKVQALR	497		
Db	413	QEDHVSMTAFARKLGDMAASNTHAILAEILAEATQGDIDLR---EPHRTSMRLQSVTEAVR	469		
Qy	498	EQFPFPLETRPLGOEIAALATHLL	521		
Db	470	RDVGHVHSDRHFSPDIAAAQVL	493		

Search completed: December 4, 2005, 09:22:26
Job time : 190 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:16:11 ; Search time 41 Seconds
(without alignments)
1234.391 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPKPAVELDRHI.....RPLQGEIATLATHLLQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	728.5	27.3	508	1	UFBSSHs
2	712	26.7	511	2	F83897
3	711.5	26.7	524	2	P75610
4	708.5	26.6	511	2	E82228
5	706	26.5	516	1	JC1172
6	695	26.0	514	2	E98245
7	695	26.0	514	2	E33040
8	694.5	26.0	657	2	S43415
9	688.5	25.8	511	2	H89758
10	679	25.4	677	2	T22333
11	677	25.4	509	2	C87368
12	670	25.1	511	2	G95945
13	666.5	25.0	657	2	A36087
14	662.5	24.8	657	2	A46128
15	657	24.6	506	2	A80596
16	652	24.4	509	2	G83009
17	643	24.1	510	1	A35251
18	620.5	23.3	510	2	B83009
19	615.5	23.1	510	2	A60488
20	603.5	22.6	461	2	H84276
21	563	21.1	500	2	B98260
22	563	21.1	500	2	A30204
23	554.5	20.8	710	2	S66313
24	543	20.4	705	2	JC5873
25	540	20.2	590	2	S70916
26	531.5	19.9	549	2	T05970
27	530	19.9	708	2	T14295
28	529	19.8	754	2	T09777
29	525	19.7	700	2	T06545

30	524	19.6	715	2	S66343	phenylalanine ammo
31	523.5	19.6	715	2	JQ2285	phenylalanine ammo
32	523	19.6	710	2	JC5872	phenylalanine ammo
33	519	19.5	725	2	S17444	phenylalanine ammo
34	516	19.3	718	2	S48726	phenylalanine ammo
35	515.5	19.3	716	2	S48725	phenylalanine ammo
36	515	19.3	712	2	S04127	phenylalanine ammo
37	514.5	19.3	716	2	S04463	phenylalanine ammo
38	514	19.3	720	2	S21174	phenylalanine ammo
39	511	19.2	713	2	S22991	phenylalanine ammo
40	510.5	19.1	667	2	T12749	phenylalanine ammo
41	509.5	19.1	701	2	S06475	phenylalanine ammo
42	509	19.1	712	2	T03663	phenylalanine ammo
43	507.5	19.0	716	2	S28185	phenylalanine ammo
44	506	19.0	723	2	S25303	phenylalanine ammo
45	505.5	18.9	721	2	A44133	phenylalanine ammo

ALIGNMENTS

RESULT 1

UFBSSHs

histidine ammonia-lyase (EC 4.3.1.3) huth [similarity] - Bacillus subtilis

N;Alternate names: histidase huth

C;Species: Bacillus subtilis

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 09-Jul-2004

C;Accession: S18810; C69643; T47089

R;Oda, M.; Sugishita, A.; Furukawa, K.

J. Bacteriol. 170, 3199-3205, 1988

A;Title: Cloning and nucleotide sequences of histidase and regulatory genes in the Bacill

A;Reference number: S18808; MUID:88257040; PMID:2454913

A;Accession: S18810

A;Molecule type: DNA

A;Residues: 1-508 <ODA>

A;Cross-references: UNIPROT:P10944; UNIPARC:UPI00000471C6; EMBL:M20659; NID:g143074; PIDN

A;Experimental source: strain IA270

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Pario, V.; Pohl, T.M.; Portetelle,

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seror,

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: C69643

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-508 <KUN>

A;Cross-references: UNIPARC:UPI00000471C6; GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAF

A;Experimental source: strain 168

R;Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.

Microbiology 141, 337-343, 1995

A;Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contain

A;Reference number: Z24350; MUID:95219086; PMID:7704263

A;Accession: T47089

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-508 <YOS>

A;Cross-references: UNIPARC:UPI00000471C6; EMBL:D31856; NID:g603765; PIDN:BA06644.1; PII

A;Experimental source: strain BGSC1A1

C;Genetics:

A;Gene: huth

C;Function:

A;Description: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic acid (urocar

A;Pathway: histidine catabolism
C;Superfamily: Bacillus subtilis histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism
F;141-143/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;142/Modified site: dehydroalanine (Ser) #status predicted

Query Match 27.3%; Score 728.5; DB 1; Length 508;
Best Local Similarity 35.5%; Pred. No. 1.5e-41;
Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9;

QY 20 IDLDQAHAVASGGARIVL-----APPARDRCASEARLGAIVREARHVYGLTTGGPL 72
DB 2 VTLGGSLTTADVARVLFDFEEAAASEMERVKRAAVERIVDEKTYIGNTGFGKF 61

QY 73 ANRLISGENVRTLOANLVHHLASGVGVLDWTTTARAMVLARLVIAQAGCAGSEGTIARL 132
DB 62 SDVLIQKEDSAALQNLILSHACGVGPPFECVSRAMLLLRANALLKGFSGVRAELIEQL 121

QY 133 IDLLNSELAPVPSRGTVGASGDLTPLAHMVCLQGRGDFLDRDGTLDGABGLRRGLQ 192
DB 122 LAFLNKRVHPVPIPOQSGSLGASGDLAPLSHLALALIGQGEVF-FEGERNPAMTGLKKAGIQ 180

QY 193 PLDLSHEDALVNGTSAWTCIALVNAHACHHLGNWAVALTALLAECLRGTEAWAAALS 252
DB 181 PVLTYSKEGLALINGTQAMTAMGVAVIEAEKAYQTERIASLTIEGLQGIIDAFEDIH 240

QY 253 DLPHFPQDAAARLRARVDSARVVRHVIAERRLDAGDGTPEAGQDAYSLRCAPOVL 312
DB 241 LARGYQEQIDVAERIRFVLSDSGLTTSQ--GELRV-----QDAYSLRCPQVH 286

QY 313 GAGPDTLAWHDRVLTIELNAVTDNVPFPPGSGVPALHGNFMGQHVALTSDALATAVTVL 372
DB 287 GATWQTLGVYVKEKLEIEMNAATDNPLIFDNGD-KVISGNGFHGQPIAFAMDFLKIAISEL 345

QY 373 AGLAERQIARLTBERLNREGLPPELHRGPAGNSFGMGAOVTATALLAEMRATG-PASIS 431
DB 346 ANTAERIERLVNPLN-DLPPFLSPHP-GLQSGAMTMQYAAASLVSENKTLAHPASVDS 403

QY 432 ISTNANQDVVSLGTIARLCREKIDRWAEILALCLAAAEELRCGSLDGVSPAGKK 491
DB 404 IPSANQEDHVSMTIARHAYQIIQNVNVLALIELICMVAQVDIR---GREKVASFTKKILEKG 460

QY 492 LVQALRQFPFLETRPLGQIEIALATHLQOS 524
DB 461 LFQEMRKVPSIQDRVFSYDIERLTDWLKES 493

RESULT 2
F83897
histidase (histidine ammonia-lyase) huth [imported] - Bacillus halodurans (strain C 125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: F83897
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83897
A;Status: preliminary
A;Molecule type: DNA
A;Cross-references: UNIPROT:Q9KBE6; UNIPARC:UPI000012CE89; GB:BA000004. NID
A;Experimental source: strain C-125
C;Genetics:
C;Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 26.7%; Score 712; DB 2; Length 511;
Best Local Similarity 33.9%; Pred. No. 1.9e-40;
Matches 171; Conservative 102; Mismatches 209; Indels 22; Gaps 8;

QY 18 RHIDLDQAHAVASGGARIVLAPPARDRCASEARLGAIVREARHVYGLTTGGFPLANRLI 77

DB 10 RSLSLHDLHRIIYEGETVGASDESMEKVKQSRKAVQIIVADEKIIYGIITTFGPKFSDFI 69

QY 78 SGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARLVIAQAGCAGSEGTIARLIDL 137

DB 70 DPDDVENLQHNLIYSHACGVGSPFETVSRMTLVLRANALLKGFSGVRLPIERLLALVN 129

QY 138 SELAPVPSRGTVGASGDLTPLAHMVCLQGRGDFLDRDGTLDGABGLRRGLQLDLS 197

DB 130 ANIHPVPIPOQSGSLGASGDLAPLSHLALVLLGEGEVFK-GTKTKASFPALKEEIEPTLT 188

QY 198 HRDALALVNGTSAWTCIALVNAHACHHLGNWAVALTALLAECLRGTEAWAAALS 257

DB 189 AKEGLALINGTQAMTAMGVAVIEAEKLAFAQSEIIASLTMEGLRGIIIDAFDEQHFARGY 248

QY 258 PGQKDAARLRARVDSARVVRHVIAERRLDAGDGTPEAGQDAYSLRCAPOVLGAGFD 317

DB 249 VEQDVARRMESVLODSQLTTRO--GELRV-----QDAYSLRCPQVHGATWQ 294

QY 318 TLAWHDRVLTIELNAVTDNVPFPPGSGVPALHGNFMGQHVALTSDALATAVTVLAGLAE 377

DB 295 TLRVYVKEKLEIEMNAATDNPLIFDNGQ-KVISGNGFHGQIQLAMDFLGIAEALANISE 353

QY 378 RQIARLTBERLNREGLPPELHRGPAGNSFGMGAOVTATALLAEMRATG-PASIS 436

DB 354 RRTERLVNPLN-DLPPFLSAAP-GVQSGVMTLQYCAASLVSENKTLAHPASVDSIPSSA 411

QY 437 ANQDVVSLGTIARLCREKIDRWAEILALCLAAAEELRCGSLDGVSPAGKKLQVQAL 496

DB 412 NQEDHVSMTIGSRHAYQIIQNVNVLALIELICMVAQVDIR---GREKVASFTKKILEKG 468

QY 497 REQFPFLETRPLGQIEIALATHL 520

DB 469 REHPVVIDQDRMPAKDIERAAKWL 492

RESULT 3
F75610
histidine ammonia-lyase (EC 4.3.1.3) [similarity] - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: F75610
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: F75610
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-524 <WHI>
A;Cross-references: UNIPROT:Q9RZ06; UNIPARC:UPI000012CE8C; GB:AE001825; NID:
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0147
A;Map position: 2
C;Superfamily: Bacillus subtilis histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;139-141/Cross-link: 5-imidazolinone (Ala-Gly) #status predicted
F;140/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.7%; Score 711.5; DB 2; Length 524;
Best Local Similarity 36.8%; Pred. No. 2.1e-40;
Matches 188; Conservative 79; Mismatches 215; Indels 29; Gaps 9;

QY 16 LDRHIDLQAHAVASGGARIVLAPPARDRCASEARLGAIVREARHVYGLTTGGFPLANR 75

DB 3 LDRDLNLEQFISVVRHGEQVELSAAARERARARTVIEQIVGEGDTPYGVNTGFGKFENV 62

QY 76 LLSGENVRTLOANLVHHLASGVGVLDWTTTARAMVLARLVIAQAGCAGSEGTIARLIDL 135

DB 63 QIDRSQALQHNLIIVSHAIGMEPLFAEVVRGMLLRAQSLSLGHSVVRVEVVELLLAL 122

Qy 136 LNSLAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRRGRLOPLD 195
Db 123 LNDALPVPVPSQGVGASGDLPLAHLALGLIGLD-IEYQGVPRPADVLAEGLSPVQ 181
Qy 196 LSHRDALVNGTSAMTGIALVNAHACHRLGNWAVALTALAECLRGRTAEWAALSCLR 255
Db 182 LOAKEGLALINGTQMLGSLALALHDAQVLLGTANLAAAMTVEARYSGSHRPFQDVVGLR 241
Qy 256 PHPGQKDAARLARVDGSAVRVHVIAERLADGDIETPEAQDAYSLRCAPQVLGAG 315
Db 242 PHPGALAVAAELREPLAGSETAPSHLTGDGVK-----QDAYSLRAVPEVHGAT 289
Qy 316 FDTLADHRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGSHVALTSDALATATVTLAG 374
Db 290 WDLAQAEVLAVERFASVTDNPLIFPETGEV--VSGGNFHPQPLAVTIDALKVAELGS 347
Qy 375 LAERQIARLTDERLNRGLPPLFHRGPPAGLNSGFMGAQVATATALLAEMRA-TGPASIHIS 433
Db 348 ISERTTEQLNPALS-GLPAFLTPN-GGLNSGFMIAQVTSAAVSENKVLSPASVSIP 405
Qy 434 TNAANQDVVSLGTAAARLCREKIDRWAEILAILALCLQAQ---AELRCGSLDGVSPAGK 490
Db 406 TSANQEDHVSMAHAARQLRQIVANVTLSIELLCAQGLDFQOLRAGRQVQAA----- 460
Qy 491 KLVALREQRPLETDRPLGQIEIAALATHLL 521
Db 461 --YEVRTFVTLTREDYFRFDLRLRGELV 489

RESULT 4
E82228
histidine ammonia-lyase (EC 4.3.1.3) [similarity] - Vibrio cholerae (strain N16961 serog
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C/Accession: E82228
R/Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, J.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellen, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: E82228
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-511 <HEI>
A/Cross-references: UNIPROT:Q9K5Q4; UNIPARC:UPI000012CE9E; GB:AE004200; GB:AE003852; NID
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC1202
A/Map position: 1
C/Superfamily: Bacillus subtilis histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/143-145/Cross-link: 5-imidazolone (Ala-Gly) #status predicted
F/144/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.6%; Score 708.5; DB 2; Length 511;
Best Local Similarity 37.4%; Pred. No. 3.2e-40;
Matches 189; Conservative 79; Mismatches 214; Indels 23; Gaps 9;

Qy 20 IDLQAHAVASGGARIVLAPPDRCRASEARLGAIVREARHVYGLTTGFGPLANRLISG 79
Db 11 LSLKQLQKVSFVLSLDLPPAIPAEASQVSEGVISGVTVGINTGFLANTKIAP 70
Qy 80 ENVTTLQANLHHLASGVPLDWTTRAMVLAELVLSIAQASGASGEGTARLIDLNSE 139
Db 71 QDLETQLKSVLSHAAGIGELMSETVRLMWLLKINSIARGSGIRLEVIQALIELVNNQ 130
Qy 140 LAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRRGRLOPLDLSHR 199
Db 131 IYPCVPKKGVSASGDLAPLHMSTVLLGEG-ARYNGKIISGLEAMKIAGLEITTLAPK 189
Qy 200 DALALVNGTSAMTGIALVNAHACHRLGNWAVALTALAECLRGRTAEWAALSCLRPHPG 259

Db 190 EGLALLNGTQASTAFALGLEFLVAEDLFASATVCGAMSVAAALGSRFPDPRIHVRGHR 249
Qy 260 QXDAARLARVDGSAVRVHVIAERLADGDIETPEAQDAYSLRCAPQVLGAGPDTL 319
Db 250 QMDAATAYRHLLDVSSSI-----QSHSNCEKVDQPYSLRCQPVQMGACLOOI 297
Qy 320 AWHDRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGSHVALTSDALATATVTLAGLAER 378
Db 298 RSAAEVLEVEANSVDNPLVFAEDGDI--ISGGNFHAEPPVMAADNLALAEIGLSUSER 355
Qy 379 QIARLTDERLNRGLPPLFHRGPPAGLNSGFMGAQVATATALLAEMRATG-PASIHISISTNAA 437
Db 356 RMALLIDSALSK-LPPLVDN-GGVNSGFMIAQVTAALASENKTLAHPASVDSIPTSAN 413
Qy 438 NQDVVSLGTAAARLCREKIDRWAEILAILALCLQAQAEELRCGSLDGVSPAGKLVQALR 497
Db 414 QEDHVSMAATFAARRLDWNGENTRGILAVEYLAQAQGLDFRAPL---KSSPRIEBEARQLR 470
Qy 498 EQPPELTDRPLGQIEIAALATHLLQ 522
Db 471 EKVPFYDKDRYFAPDIEK-ANALLQ 494

RESULT 5
JC1172
histidine ammonia-lyase (EC 4.3.1.3) [validated] - Streptomyces griseus
N/Alternate names: histidase
C/Species: Streptomyces griseus
C/Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 09-Jul-2004
C/Accession: JC1172; PC1104; B42299
R/Wu, P.C.; Kroening, T.A.; White, P.J.; Kendrick, K.E.
Gene 115, 19-25, 1992
A/Title: Histidine ammonia-lyase from Streptomyces griseus.
A/Reference number: JC1172; MUID:92307427; PMID:1612436
A/Accession: JC1172
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-516 <WUP>
A/Cross-references: UNIPROT:P24221; UNIPARC:UPI000000471C7
A/Experimental source: strain NRRL B-2682
A/Accession: PC1104
A/Molecule type: protein
A/Residues: 1-12 <WU2>
A/Cross-references: UNIPARC:UPI0000173007
R/Wu, P.C.; Kroening, T.A.; White, P.J.; Kendrick, K.E.
J. Bacteriol. 174, 1647-1655, 1992
A/Title: Purification of histidase from Streptomyces griseus and nucleotide sequence of t
A/Reference number: A42299; MUID:92165741; PMID:1537807
A/Accession: B42299
A/Molecule type: DNA
A/Residues: 1-516 <WUA>
A/Cross-references: UNIPARC:UPI00000471C7; GB:M77841; NID:GL53316; PIDN:AAA26769.1; PID:5
C/Genetics:
A/Gene: huth
C/Function:
A/Description: EC 4.3.1.3; histidine ammonia-lyase; catalyzes the formation of (E)-3-(1H-
A/Pathway: histidine catabolism
A/Note: specific for L-histidine, shows no activity against D-histidine
A/Note: histidinol phosphate functions as competitive inhibitor; high concentrations of t
as in other histidases, dehydroalanine plays an important role in catalysis
C/Superfamily: Bacillus subtilis histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F/145-147/Cross-link: 5-imidazolone (Cys-Gly) #status predicted
F/146/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.5%; Score 706; DB 1; Length 516;
Best Local Similarity 38.6%; Pred. No. 4.8e-40;
Matches 192; Conservative 69; Mismatches 212; Indels 24; Gaps 8;

Qy 27 AVASGGARIVLAPPDRCRASEARLGAIVREARHVYGLTTGFGPLANRLISGENVRLQ 86
Db 20 AVARHGAVELSAVAEALAAARLIVDALEAKPEPVYGVSTGFGALASRHIGTELRLAQ 79

```
QY 87 ANLVHHLASGVGVPLDWTMTARVMVRLVSIQAQASGSEGTIARLIDILNSELAPAVPS 146
Db 80 RNIIVRSHAAAGMPREVERVVALMFLRLKTVASGHTGVRPEVAQTMADVLNAGITPVVHE 139
QY 147 RGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTDLRDGAEGLRRGELQPLDLSHRDALAVN 206
Db 140 YSLGCGSGDLAPLSHCHALTLMGGEAEGPGDGTVRPAGELLAHHGIAFVELREKEGLALN 199
QY 207 GTSAMTGTIALVNAHACHRLGNWVALTALLAECLRGRTAEWAALSLDLRPHPGQKDAAR 266
Db 200 GTDGMGLVMALADLNLITSADITAALESALLGTDKVLAPELHAIIRHPHQQGVSAADN 259
QY 267 LRARVDSGARVRVHVIARERLDAGDICTEPEAGODAYSLRCAPQVLGAGFDTTLAWHDRVL 326
Db 260 MSKVLASGSLTGHH-----QODAPRV-----QDAYSVRCAPQVNGAGRDTLDHAALVA 307
QY 327 TIENAVTDNPVPPDGVSVALHGGNPMGOHVALTSDALATATVTLVGLAERQIAKLTDE 386
Db 308 GRELASVDNPNVPLPDGRVES--NGNFHGAPVAVYVLDFLAIVAADLGSICERETDRLLDK 365
QY 387 RLNRGLPPFLHRGPAGLNSGFMGAQVATATALLAEM-RATGPASIHSTISNAANQDVVSLG 445
Db 366 NRSHGLPPFL-ADDAGVDSGLMTAQYTOAALVSEMKEKLAVPASADSLPSSAMQEDHVSIG 424
QY 446 TIAARLCREKIDRWAEITLALCLQAELRCGSLDGVSPAGKLVQALR-----EQFP 501
Db 425 WSAARKLRATVDNLARIVAVELVAAATRAIELRAAEGLT-PAPASEAVVAALRAAGAEGPG 483
QY 502 PLETDRLPGOEIAALAT 518
Db 484 P---DRFLAPDLAAADT 497

RESULT 6
AG3040
hiscidine ammonia-lyase huth [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AG3040
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; W. O., L.
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutysavin, T.; Levy, R.; Li, M.; Mclell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-K. mm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG3040
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>
A:Cross-references: UNIPROT:Q8U827; UNIPARC:UPI000012CE88; GB:AE008689; PIDN:AAL447.1.1;
C:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: huth
A:Map position: linear chromosome
C:Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 26.0%; Score 695; DB 2; Length 514;
Best Local Similarity 38.1%; Pred. No. 2.7e-39;
Matches 179; Conservative 66; Mismatches 203; Indels 22; Gaps 8;

QY 50 ARLGAVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVPLDWTMTARAM 109
Db 43 ARIAETAAGNAPVYGINTEGFKLASIKIDAADVATLQRNLILSHCCGVGVPLPENVRLLI 102
QY 110 VLARLVSIQAQASGSEGTIARLID-LLNSELAPAVPSRGTVCASGDLTPLAHMVLCLQG 168
Db 103 MALKVLSLGRGASGV-RLELVRLIEAMLEKGVIPVPEKGSVGSASGDLAPLAHMAAVMMG 161
QY 169 RGFDRDGTDLGAEGLRGRLQPLDLSHRDALVNGTISAMTGTIALVNAHACHRLGNW 228
Db 162 EGEAF-YEGAPLPAGEALAKAGLTPVVLAAKGLALINGTQTSTALALAGLFRHRAAQA 220
QY 229 AVALTALLAECLRGRTAEWAALSLDLRPHPGQKDAARLARVDSGARVRVHVIARERLD 288
Db 221 ALITGALSTDAAMGSSAPFPHDIHSLRGHKQIDAGALRNLLGSEIRVSHIEGDERV- 279
QY 289 AGDIGTEPEAGODAYSLRCAPQVLGAGFDTTLAWHDRVLTIELNAVTDNPVPPDGSGVPAL 348
Db 280 -----QDPYCIRCPQVDGACLDLLRQVARTLEIANAVTDNPLVLSDNSV--V 326
QY 349 HGGNFMGOHVALTSDALATATVTLVGLAERQIAKLTDELRNRLGPPFLHRGPAGLNSGFM 408
Db 327 SGNFHAEPVAFADQTLAICEIGAIAQRRVALLVDPALSYGLPAFLSKKP-GLNSGLM 385
QY 409 GAQVTTALLAEWRATG-PASIHSTISNAANQDVVSLGTIAARLCREKIDRWAEIILAILA 467
Db 386 IAEVTSAAALMENKQWHPASVDSTPSTANQEDHVSMAHGARRLLPMTDNLPAIIGIEA 445
QY 468 LCLQAELRCGSLDGVSPAGKLVQALRQFPPLPTEDRPLGQOEIAALA 517
Db 446 LSAVQGVELR---GPLKTSFELQKVIAVLRVVVPVPSLEEDRYMAPDLKAAA 492
```

RESULT 7

```
E98245
huth gene homolog [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: E98245
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98245
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>
A:Cross-references: UNIPROT:Q8U827; UNIPARC:UPI000012CE88; GB:AE007870; PIDN:AAK89487.1;
C:Genetics:
A:Gene: AGR_L1825
A:Map position: linear chromosome
C:Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 26.0%; Score 695; DB 2; Length 514;
Best Local Similarity 38.1%; Pred. No. 2.7e-39;
Matches 179; Conservative 66; Mismatches 203; Indels 22; Gaps 8;

QY 50 ARLGAVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVPLDWTMTARAM 109
Db 43 ARIAETAAGNAPVYGINTEGFKLASIKIDAADVATLQRNLILSHCCGVGVPLPENVRLLI 102
QY 110 VLARLVSIQAQASGSEGTIARLID-LLNSELAPAVPSRGTVCASGDLTPLAHMVLCLQG 168
Db 103 MALKVLSLGRGASGV-RLELVRLIEAMLEKGVIPVPEKGSVGSASGDLAPLAHMAAVMMG 161
QY 169 RGFDRDGTDLGAEGLRGRLQPLDLSHRDALVNGTISAMTGTIALVNAHACHRLGNW 228
Db 162 EGEAF-YEGAPLPAGEALAKAGLTPVVLAAKGLALINGTQTSTALALAGLFRHRAAQA 220
QY 229 AVALTALLAECLRGRTAEWAALSLDLRPHPGQKDAARLARVDSGARVRVHVIARERLD 288
Db 221 ALITGALSTDAAMGSSAPFPHDIHSLRGHKQIDAGALRNLLGSEIRVSHIEGDERV- 279
QY 289 AGDIGTEPEAGODAYSLRCAPQVLGAGFDTTLAWHDRVLTIELNAVTDNPVPPDGSGVPAL 348
Db 280 -----QDPYCIRCPQVDGACLDLLRQVARTLEIANAVTDNPLVLSDNSV--V 326
QY 349 HGGNFMGOHVALTSDALATATVTLVGLAERQIAKLTDELRNRLGPPFLHRGPAGLNSGFM 408
Db 327 SGNFHAEPVAFADQTLAICEIGAIAQRRVALLVDPALSYGLPAFLSKKP-GLNSGLM 385
QY 409 GAQVTTALLAEWRATG-PASIHSTISNAANQDVVSLGTIAARLCREKIDRWAEIILAILA 467
Db 386 IAEVTSAAALMENKQWHPASVDSTPSTANQEDHVSMAHGARRLLPMTDNLPAIIGIEA 445
```

Db 386 IAEVTSAAAMSENKQMAHPASVDSTPTTSANOEDHVSMAHGARRLLPMTDNLFAILGIEA 445

Qy 468 LCLAQAAELRCGSLDGVSAGKLVQALRRQFPFPLETDRPLRQEI AALA 517

Db 446 LSAVQGVFLR---GPLTKSPELOKVIARLVVVPSLEEDRYMAPDLKAAA 492

RESULT 8

S43415

histidine ammonia-lyase (EC 4.3.1.3) - human

N;Alternate names: histidase

C;Species: Homo sapiens (man)

C;Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S43415

R;Suchi, M.; Harada, N.; Wada, Y.; Takagi, Y.

Biochim. Biophys. Acta 1216, 293-295, 1993

A;Title: Molecular cloning of a cDNA encoding human histidase.

A;Reference number: S43415; MUID:94060103; PMID:7916645

A;Accession: S43415

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-657 <SUC>

A;Cross-references: UNIPROT:P42357; UNIPARC:UPI000012CE8E; DDBJ:D16626; NID:g451209; PID

C;Genetics:

A;Gene: GDB:HAL; HIS

A;Cross-references: GDB:120746; OMIM:235800

A;Map position: 12q22-12q23

C;Function:

A;Description: catalyzes the formation of (E)-3-(1H-imidazol-4-yl)-propenoic acid (iroce

A;Pathway: histidine catabolism

C;Superfamily: histidine ammonia-lyase

C;Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism

F;253-255/Cross-link: 5-imidazolone (Ala-Gly) #status predicted

F;254/Modified site: dehydroalanine (Ser) #status predicted

Query Match 26.08; Score 694.5; DB 2; Length 657;

Best Local Similarity 35.78; Pred. No. 3.9e-39;

Matches 193; Conservative 81; Mismatches 227; Indels 39; Gaps 11;

Qy 6 AMSDFIPSPQEGVLYSKYREPEKYLTDGRLTTEDLVNLGKRYKILTPTAERVRQ 147

Db 88 AMSDFIPSPQEGVLYSKYREPEKYLTDGRLTTEDLVNLGKRYKILTPTAERVRQ 147

Qy 47 ASEARLGAIVREARHVYGLTGFPLANRLISGENVRLQANLVHLSAGVGPVLDWTTA 106

Db 148 KSREVIDSIKEKTVVYGITTFGKFARTVPIPKLQELQVNLVRSHSGVGKPLSPERC 207

Qy 107 RAMVLARLVSTAOAGSASECTIARLDLNSLAPAPVSRGTGASGDLTPLAHMWLCL 166

Db 208 RMLLALRLNLAKYSGISLETLKQVEMFNASCCLPYPEKGTGASGDLAPLUSHLAGL 267

Qy 167 QGRGDFLDRDCTRLDGAELRRGRQLPDLDSHRDALALVNGTSAMTGTALVNAHACRHLG 226

Db 268 VGEKWKSPKSGWADAKVLEHGLKPVILKPKGLALINGTQMTLSLGCENVERASAIA 327

Qy 227 NWAVALTALLAECLURGTEAWAALS DLRPHPGQKDAARLARVDSGARVVRHVIAERR 286

Db 328 ROADIVAALTLEVLKGTTKAPDTHALRPHRGQIEVAFRPSLSDSDHHPSEIAESHRF 387

Qy 287 LDAGDIGTEPAGODASLRCAPOVLGAGPTTLAHHDRVLTIENAVTDNP-VFPPPGSV 345

Db 388 CD-----RVQDAYTLRCCPQVGHVNDTIAFVKNIIITELNSATDNPVFANRGE- 437

Qy 346 PALHGGNFMGQHVALTSDALATATVTLAGLAERQIARLTDERLNRGLPPLHRRGPAGLNS 405

Db 438 -TVSGGNFHEGYPKALDYLAIGHLELAISERIERLCNPPLSE-LPAFL-VAEGGLNS 494

Qy 406 GFMGCAQVTTALLAEMRA-TGPASIHSTISNAANQDVVSLGTIAARLCREKIDRWAEILA 464

Db 495 GFMLAHCTAAALVSENKALCHPSSVDSLSSTAATEDHVSNGMAARALRVIEHVEQVLA 554

Qy 465 ILALCLQAABE-LRCGSLDGVSAGKLVQALRRQFPFPLETDRPLRQEI AALHLLQ 523

Db 555 IELLAACQGIETFLR---PLKTTTPL-EKVVLYRSVVVRPWIKDRFMAPDIEAAHRLLEQ 610

RESULT 9

H89758

histidine ammonia-lyase [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C;Accession: H89758

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shibata, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: H89758; MUID:21311952; PMID:11418146

A;Accession: H89758

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-511 <KUR>

A;Cross-references: UNIPARC:UPI0000165A2C; GB:BA000018; PID:g13699925; PIDN:BA041224.1; C

C;Genetics:

A;Gene: huth

C;Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 25.84; Score 688.5; DB 2; Length 511;

Best Local Similarity 35.44; Pred. No. 7.2e-39;

Matches 174; Conservative 89; Mismatches 208; Indels 21; Gaps 7;

Qy 33 ARIVLAPPDRDCRSEARLGAIVREARHVYGLTGFPLANRLISGENVRLQANLVH 92

Db 30 SKIEIIDDALERVKKSRVAVRIENETVYGITTFGLFSDVRIDPTQYNELNLI 89

Qy 93 LAGVGVPVLDWTARAVMLARLVSTAOAGSASECTIARLDLNSLAPAPVSRGTGVA 152

Db 90 HAGLGEPPSPKSEVALVMMILRLNTLLKHSGLTLEVRQLQFFINERIIPIPOGSLGA 149

Qy 153 SGLTPLAHMWLCLQGRGDFLDRDCTRLDGAELRRGRQLPDLDSHRDALALVNGTSAMT 212

Db 150 SGLDAPLPLSHLALALIGEGKLYLR-GEKQSDVDVLRELNRQPLNLQAKEGLALINGTQAMT 208

Qy 213 GIALVNAHACRHGNNAVALTALLAECLRCRTAWAALS DLRPHPGQKDAARLARVD 272

Db 209 AQGVSYIEADLGYQSEWTAALTHQSLNGIIDIAYRHDVHSVRNFQEQINVAARMRWLE 268

Qy 273 GSARVVRHVIAERRLDAGDIGTEPAGQDAYSRLCAPQVLGAGPTTLAHHDRVLTIENLA 332

Db 269 GSTLTTRQ--AEIRV-----QDAYTLRCIPQIHGASQVFNYYKQLEFEMNA 314

Qy 333 VTQNP-VFPPDGSVPALHGNFMGQHVALTSDALATATVTLAGLAERQIARLTDERLNRG 391

Db 315 ANDNPLIFEBANETFTVSGNFHQPTAFALDHLKLGVELANVSERRLERLVNPLQNGD 374

Qy 392 LPPLHRRGPAGLNSGFMAQVTTALLAEMRATG-PASIHSTISNAANQDVVSLGTIAAR 450

Db 375 LPAPLSEPE-GLOSGAMIMQYAAASLVSENKTLAHPASVDSITSSANQEDHVSMTGTAAR 433

Qy 451 LCREKIDRWAEILAILALCLQAALRCGSLDGVSAGKLVQALRRQFPFPLETDRPLG 510

Db 434 HGQIENARRVLAIECVIALQAAELK--GVGSLSPKTRKRYEERFSIIVPSIITHDRQFH 490

Qy 511 QETAAALATHLLQ 522

Db 491 KDIEAVAQYLKQ 502

RESULT 10

T22333

histidine ammonia-lyase (EC 4.3.1.3) - Caenorhabditis elegans

N;Alternate names: histidase; protein F47B10.2

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T22333

R;Harris, B.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19550
A;Accession: T22333
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-677 <WIL>
A;Cross-references: UNIPROT:Q20502; UNIPARC:UPI000012CE8A; EMBL:Z68004; PIDN:CAA919/2.1;
A;Experimental source: clone F47B10
C;Genetics:
A;Gene: CESP:F47B10.2
A;Map position: X
A;Introns: 24/3; 77/1; 111/2; 178/1; 255/1; 335/1; 393/1; 523/1; 544/3; 604/2
C;Superfamily: histidine ammonia-lyase
C;Keywords: ammonia-lyase; carbon-nitrogen lyase
F;269-271/Cross-link: 5-imidazolone (Cys-Gly) #status predicted
F;270/Modified site: dehydroalanine (Ser) #status predicted

Query Match 25.4%; Score 679; DB 2; Length 677;
Best Local Similarity 35.3%; Pred. No. 4.5e-38;
Matches 182; Conservative 85; Mismatches 231; Indels 18; Gaps 10;

QY 3 PMLAMSPKPAVELDRHIDL-DQAHAVASGGAIVLAPPARDRCRASEARLGAIVREARH 61
DB 120 PMV-LAPPTTKLLLDGNLSLPEDLVRCCKEQAQLSMESEDIRKARTFLEKIASHRA 178
QY 62 VYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGPVLDWTTARAWLARLVLSIAOGA 121
DB 179 VYGVTTGFGTFSNVTPPEKKLQLNLRSHATGGEPLAPNARWLLALRNILIAKH 238
QY 122 SGASEGTIARLLDLLNSELAPVPSRTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRLD 181
DB 239 SGISVENIKKMAAFNAFCVSYVPQGTGCSGDLCPLAHLALGLLGECKWSPPTTGWQP 298
QY 182 GAGELRCRLOPLDLSHRDALVNGTSAMTALVNAHACRHLGNWAVALTALLAECLR 241
DB 299 ADVLKNKNIPELPGKEGLALINGTQMTVALGAYTLERAHNIARQADVIAALSVDLK 358
QY 242 GRTEAWAALSDLRPHGQKAAARLARVDGSRVVRHVIAERLLDAGDIGTEPEAGOD 301
DB 359 GTTRAVDPDHRTPRHGQNLRLALLHSEANSQ--IAESHRNCTKV-----QD 409
QY 302 AYSLRCAPOVLGAGFTLAWHDHRLVLTIELNAVTDNVPFPPDGSPALHGNFMGQHVALT 361
DB 410 AYTLRCVPQVGVVHDTIEPVREIITMNSATDNPLVPADRE-EIISGNGPHGEYPAKA 468
QY 362 SDALATAVTVLAGLAERQIARLDRLNRGLPPLPHRGAGLNSGFMGAQVTTATALLAEM 421
DB 469 LDFLATAVAEFAQMSERRLERLVNKELS-GLPTFL-TPDGGLNSGFMVTQLCAASLVSEN 526
QY 422 RA-TGPASIHSTNAANQDWSLGTIAARLCREKIDRWAEIILAILALCLAAQAEELRCGS 480
DB 527 KVLCHSSVDSIPTSCNQEDHVMGFGFAARKALTVVHEVAVLAMELLAACQIEFL--K 584
QY 481 GLDGVPAGKKLQVALREQPPPLETRPLGQETAAAL 516
DB 585 PLISTAPL-HKIQLVRSVAPPLNEDRYMKPEIDAV 619

RESULT 11
C87368
histidine ammonia-lyase [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: C87368
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Jol
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87368
A;Status: preliminary

A;Molecule type: DNA
A;Residues: 1-509 <STO>
A;Cross-references: UNIPROT:P58082; UNIPARC:UPI000012CE8B; GB:AE005673; NID:gl3422237; P1
C;Genetics:
A;Gene: CC0959
C;Superfamily: Bacillus subtilis histidine ammonia-lyase

Query Match 25.4%; Score 677; DB 2; Length 509;
Best Local Similarity 37.9%; Pred. No. 4.3e-38;
Matches 187; Conservative 61; Mismatches 215; Indels 30; Gaps 9;

QY 32 GARIVLAPPARDRCRASEARLGAIVREARHVYGLTTGFGPLANRLISGENVRTLQANLVH 91
DB 22 GASARLAESAWPVIABSAVAQVRIKAGPEVYGVYNTGFGKLASVRIGDADLETQRNVL 81
QY 92 HLASGVGPVLDWTTARAWLARLVLSIAOGASGSECTIARLLDLLNSELAPVPSRTVG 151
DB 82 SHAAAGVGPSPVPVIRLMMALKLASIAOGASGVVRVETVRMLEEMLVEGLTPVVPVCGSVG 141
QY 152 ASGDLTPLAHMVLCLQGRGDFLDRDGTRLDGAELRRGRRLQPLDLSHRDALALVNGTSAM 211
DB 142 ASGDLAPLSHMAATMIGVEIF-VGQRLPAAQALAAQAGLEPLTLGPKGLALLNGTQPS 200
QY 212 TGIALVNAHACRHLGNWAVALTALLAECLRGRTEAWAALSDDLPHPGQKDAARLARV 271
DB 201 TANALAGLFEAERLFQSAVLVTGALSTEAAGSGDTPDPRIHTLRRHVQGIETAAALRALM 260
QY 272 DGSARVVRHVIAERLLDAGDIGTEPEAGODAYSLRCAPOVLGAGFTTLLAWHDHRLVLTIELN 331
DB 261 SASEIRASHL-----KEDERVQDPVCLRCQPQVMGAALDILROAATTTATEAN 308
QY 332 AVTDNVPFPPDGSPALHGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERLNRG 391
DB 309 CVSDNPLIFPEAD-EALSGNFHAFSPVAFADMIALAVCEIGSIAERRIAMLVDPALSG 366
QY 392 LPPFLHRCGAGLNSGFMGAQVTTATALLAEMRATG-PASIHSTNAANQDVSLSGTIAAR 450
DB 367 LPAFLTPKP-GLNSGFMIPQVTAALVSENKQRAVPASVDSIPTSANQEDHVSMAAHGAR 425
QY 451 LCREKIDRWAEIILAILALCLAAQ-----AELRCGSLDGVSPPAGKKLQVAL-REQPPLET 505
DB 426 RLLAMVENADAVLGIELLAAQCGDFHAPLRSAALEA-----VRALTRSKVPHLSD 477
QY 506 DRPLGQETIAALAT 518
DB 478 DRHFHPDMEAAVT 490

RESULT 12
G95945
probable histidine ammonia-lyase (EC 4.3.1.3) [imported] - Sinorhizobium meliloti (strain
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C;Accession: G95945
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: G95945
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-511 <KUR>
A;Cross-references: UNIPROT:O31197; UNIPARC:UPI000012CE93; GB:AL591985; PIDN:CAC49231.1;
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation

C:Genetics:

A:Gene: hucH1, SMD211165

A:Genome: plasmid

C:Superfamily: Bacillus subtilis histidine ammonia-lyase

C:Keywords: ammonia-lyase; carbon-nitrogen lyase

Query Match 25.1%; Score 670; DB 2; Length 511;

Best Local Similarity 35.8%; Pred. No. 1.3e-37;

Matches 172; Conservative 72; Mismatches 202; Indels 22; Gaps 8;

Qy 50 ARLGAVREARHVVGLTGFGLANRLISGENVRLTQANLVHLLASGVGPVLDWTTARAM 109

Db 40 ARIAEIVAGNAVYVINGTGFKLASIKIDSDVATLQNLILSHCCGCGPLTDIVRLI 99

Qy 110 VLARLVIAQAGSASGTTIARLID-LNSELAPAVPSRGTVGASGDLTFLAHMVLCLQG 168

Db 100 MALKLILSGRGASGV-RLELVRLTEAMDKGVIPLIPEKGSVGSAGDLAPLAHMAAVMMG 158

Qy 169 RGDPLDRDGLDGAELRGRLQPLDLSHRDALALVNGTSMTGIALVNAHACRHLGNW 228

Db 159 HGEAFFA-GERMKGDAAKAGSPVTLAAKEGLALINGTQVSTALAGLFRHRAQQA 217

Qy 229 AVALTALLAECLRGRTAWAAALSDLRPHPGQKDAARLARARVDGSARVVRHVIAERRLD 288

Db 218 ALITGALSTDAWMGSSAPFPDHTLRHGKQIDTAALQQLTGSPIRQSHIEGDERV- 276

Qy 289 AGDIGTPEAGQADAYSURCAQVLAGFDTLAWHDRVLTIELNAVTDNVPFPGSVPAL 348

Db 277 -----QDPYCIRCQPOVDGACLDLRSVAATLTIEANAVTDNPLVLSNSV--V 323

Qy 349 HGGNFMGQHVLTSDALATVTVLAGLAERQIARLTDERLNRGLPPLHRRGPAGLNSGFM 408

Db 324 SGGNFHAEPVAFADQALAVCVSIGATSORRIALLVDPALSYGLPAPFAKKP-GLNSGLM 382

Qy 409 GAQVTAALLAE-WRATGPASIHISSTNAANQDVVSLGTIAARLCREKIDRWAEIILAILA 467

Db 383 IAEVTSAAIENKQLSHPASVSTPISANQEDHVSNAACHARRLLQMTENLFSITGIEA 442

Qy 468 LCLQAALRCGSLDGVSPAGKLVQALRQFPFPLETRPLQOEIAA 515

Db 443 LAAVQGIETFRAPL---TTPSELKAAAARVGVSSIEEDRYMADDLKA 487

RESULT 13

A36087

histidine ammonia-lyase (EC 4.3.1.3) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 25-Jan-1991 #sequence_revision 25-Jan-1991 #text_change 09-Jul-2004

C:Accession: A36087

R;Taylor, R.G.; Lambert, M.A.; Sexsmith, E.; Sadler, S.J.; Ray, P.N.; Mahuran, D.J.; McI

J. Biol. Chem. 265, 18192-18199, 1990

A;Title: Cloning and expression of rat histidase. Homology to two bacterial histidases a

A;Reference number: A36087; MUID:91009306; PMID:2120224

A;Accession: A36087

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-657 <TAY>

A;Cross-references: UNIPROT:P21213; UNIPARC:UPI00000471C3; GB:M58308; GB:J05653; NIJ:G20

C:Superfamily: histidine ammonia-lyase

C:Keywords: ammonia-lyase; carbon-nitrogen lyase; histidine catabolism

P;253-255/Cross-link: 5-imidazolone (Ala-Gly) #status predicted

F;254/Modified site: dehydroalanine (Ser) #status predicted

Query Match 25.0%; Score 666.5; DB 2; Length 657;

Best Local Similarity 35.7%; Pred. No. 3e-37;

Matches 185; Conservative 81; Mismatches 231; Indels 21; Gaps 10;

Qy 10 PKPAVELD-RHIDLDQHAVASGARIVLAPPARDRCRASEARLGVAREARHVVGLTGG 68

Db 110 PEKYIALDGSLSLSTEDLVNLGKGYIKLTSIAEKKVQQSREVIDSIKERTVYVGIITG 169

Qy 69 FGPLANRLISGENVRLTQANLVHLLASGVGPVLDWTTARAVLARLVIAQAGSASEGT 128

Db 129 IARLIDLNSELAPAVPSRGTVGASGDLTFLAHMVLCLQGRGDFLDRDGTFLDGAELRR 188

QY 189 GRLQPLDLSHRDALVNGTSAMTGIALVNAHACRHLGNWAVALTALLAECLRGTRTEAWA 248
Db 290 HGLKPIVLKPEGLALINGTOMITSLGCEALERASAIARQADIVAALTLEVLKGTTKAFD 349
QY 249 AALSDLRPHGQKDAARLARARVDGSGARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCA 308
Db 350 TDHAVRPHRGQTEVAFRFRSLDSDHHPSEIAESHRFCD-----RVQDAYTLRCC 400
QY 309 PQVLGAGFTLAWHDRVLTITLNAVTDNP-VFPDGSVPALHGNFMGOHVALTSDALAT 367
Db 401 PQHGVVNDTIAFVKDIITELMSATDNPVFSRGE--TISGNFHFGEYPAKALDYLA 458
QY 368 AVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRA-TGP 426
Db 459 GVHELAAISERRIERLCNPSLSL-LPAFL-VASGGLNSGFMIAHCTAAALVSESKALCHP 516
QY 427 ASHISISTNANQDVVSLGTIARLCREKIDRWAEIILAILALCIAQAE-LRCGSGLDGV 485
Db 517 SSVDSLSTSAATEDHVSIMGGAARKALRVVHEVQVLATIELLAACQGIIBFLR---PLKTT 573
QY 486 SPAGKKLVQALREQFPFPLETDRPLGQEIATALLHLLQQ 523
Db 574 TPL-EKYVDLVRVVRPWIKDRFMAPDIEAAHRLLLDQ 610

RESULT 15
AE0596
histidine ammonia-lyase (EC 4.3.1.3) [imported] - Salmonella enterica subsp. enterica serovar typhi
A:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AE0596
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhi
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AE0596
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <PAR>
A:Cross-references: UNIPARC:UPI00005A0F7; GB:AL513382; PIDN:CAD05239.1; PID:gl6502.09;
C:Gene: hutH
C:Superfamily: Bacillus subtilis histidine ammonia-lyase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase

Query Match 24.6%; Score 657; DB 2; Length 506;
Best Local Similarity 36.4%; Pred. No. 9.4e-37;
Matches 186; Conservative 76; Mismatches 215; Indels 34; Gaps 11;

QY 20 IDLDQAHAVASGARIVLAPPARDRCRASEARLGAIVREARHYGLTGTGFLANRLISG 79
Db 11 LSLQLYDVRHVPQRLDASALDGINASVACVNDIVAEGRYAGINTGFLLAQTRIAD 70
QY 80 ENVRTLOANLVHHLASGVGVLDTWTARAVLARLVSIAGGASGASEGTIARLIDLINSE 139
Db 71 EDLQNLQRLSVLASHAAGVGDPDLDAMVRLIMVLKINSLARGFSGIRLSVIEALIALVNA 130
QY 140 LAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTRLDCAEGLRRGRLOPLDLSHR 199
Db 131 VYPLIPAKGSVGASGDLAPLHLISLTLGEGK-ARWQGEWLPAQAALKKAGLEPVALA 189
QY 200 DALALVNGTSAMTGIALVNAHACRHLGNWAVALTALLAECLRGTRTEAWAALSDLRPHFG 259
Db 190 EGLALINGTOASTAFALRGLFEAQELFASAVCGALTTEAVLGSRCRPFDAIHAARGQQG 249
QY 260 OKDAAARLARVDGSGARVVRHV-----IAERRLDAGDIGTEPEAGQDAYSLRCAPOVLG 313
Db 250 QIDV-----ARLFRHLTDTSAIESHHCHKV-----QDPYSLRCQPQVNG 291

QY 314 AGFDTILAWHDRVLTITIELNAVTDNP-VFPDGSVPALHGNFMGOHVALTSDALATAVTVL 372
Db 292 ACTQLRQTKEVLLAEANAVSDNPLVFAAGEV--ISGNGFHAEPVAMAADNLALAI AEI 349
QY 373 AGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRA-TG-PASIIHS 431
Db 350 GALSERRITLMDMKHMSQ-LPPFLVKN-GGVNSGFMIAQVTAALASENKALAHPHSVDS 407
QY 432 ISTNANQDVVSLGTIARLCREKIDRWAEIILAILALCIAQAE-LRCGSGLDGVSPAGKK 491
Db 408 LPTSANQEDHVSNAAPAGRRRLWEMAANTRGVIAVEWLAACQGIDLR--EGLTS-SPLEEQ 464
QY 492 LVQALREQFPFPLETDRPLGQEIATALLHLLQQ 522
Db 465 ARQTLREVRVAHYTQDRFFAPDIECATITLLAQ 495

Search completed: December 4, 2005, 09:27:09
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:15:56 ; Search time 232 Seconds
(without alignments)
1599.603 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPRPKPAVELDRHI.....RPLQGEIATAALTHLLQQSPV 526

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	851	31.9	507	1 HUTH_SYMTH	Q67jh4 symbiobacte
2	814.5	30.5	539	2 Q8GMG0 STRGL	Q8gmgo streptomyc
3	814	30.5	508	2 QANR19 9DELT	Q4nr19 anaeromyxob
4	801	30.0	508	1 HUTH_THETN	Q8rbh4 thermomane
5	797.5	29.9	514	1 HUTH_GLOVI	Q7ncb3 gloeobacter
6	792.5	29.7	523	2 Q9KHJ9 9ACTO	Q9khj9 streptomyc
7	763	28.6	499	2 Q4H6E7 9DEIO	Q4he67 deinococcus
8	762.5	28.6	526	1 HUTH_AZOSE	Q5nrx8 azoarcus sp
9	753	28.2	510	1 HUTH_STIAU	Q93tx3 stigmatella
10	749.5	28.1	507	1 HUTH_BURVA	Q62lj6 burkholderi
11	749.5	28.1	507	1 HUTH_BURPS	Q63sh6 burkholderi
12	746	28.0	507	1 HUTH_TREDE	Q73q56 treponema d
13	733	27.5	504	1 HUTH_GEOKA	Q51310 geobacillus
14	729.5	27.3	507	2 Q4LS33 9BURK	Q4ls33 burkholderi
15	728.5	27.3	508	1 HUTH_BACSU	P10944 bacillus su
16	726.5	27.2	519	1 HUTH_BRAJA	Q89gv3 bradyrhizob
17	726	27.2	514	1 HUTH_CLOTE	Q89lq1 clostridium
18	726	27.2	524	2 Q5Z0G8 NOCPA	Q5z0g8 nocardia fa
19	726	27.2	528	2 Q8XSV2 RALSO	Q8xsv2 ralstonia s
20	725	27.1	511	1 HUTH2_FUSNN	Q8rdu4 fusobacteri
21	723.5	27.1	516	2 Q7PSN4 FUSNV	Q7psn4 streptomyc
22	721.5	27.0	512	1 HUTH_STRCO	Q9swu1 streptomyc
23	721	27.0	500	1 HUTH_BACSK	Q5waz6 bacillus cl
24	715	26.8	495	1 HUTH_THEVO	Q978n8 thermoplasm
25	715	26.8	512	1 HUTH_STRAW	Q82133 streptomyc
26	712	26.7	511	1 HUTH_BACHD	Q9kbe6 bacillus ha
27	711.5	26.7	524	1 HUTH_DESPS	Q6akp3 desulfotale
28	711.5	26.7	524	1 HUTH_DEIRA	Q9rz06 deinococcus
29	710.5	26.6	541	2 Q84FL5 ENTAG	Q84fl5 enterobacte
30	708.5	26.6	510	1 HUTH_CHRVO	Q7p188 chromobacte
31	708.5	26.6	511	1 HUTH_VIBCH	Q9ksq4 vibrio chol

32	706	26.5	496	1 HUTH_THEAC	Q9hli6 thermoplasm
33	706	26.5	514	1 HUTH_STRGR	P24221 streptomyc
34	701.5	26.3	513	1 HUTH_SHEON	Q8ekj4 shewanella
35	700	26.2	511	1 HUTH_VIBPA	Q87q77 vibrio para
36	699.5	26.2	515	1 HUTH_GLUOK	Q5frr8 gluconobact
37	698.5	26.2	510	1 HUTH_VIBF1	Q5e0c6 vibrio fisc
38	696	26.1	925	1 HUTHI_BRUME	Q8yd09 brucella me
39	695.5	26.1	657	2 Q4VB93 HUMAN	Q4vb93 homo sapien
40	695	26.0	514	1 HUTH_AGRF5	Q8u8z7 agrobacteri
41	694.5	26.0	657	1 HUTH_HUMAN	P42357 homo sapien
42	694.5	26.0	657	2 Q4VB92 HUMAN	Q4vb92 homo sapien
43	692.5	26.0	509	2 Q4IYS1 AZOVI	Q4lysl1 azotobacter
44	689	25.8	664	2 Q6GMC2 XENLA	Q6gmc2 xenopus lae
45	688.5	25.8	504	1 HUTH_STAAM	P64415 staphylococ

ALIGNMENTS

RESULT 1

ID	HUTH_SYMTH	STANDARD;	PRT;	507 AA.
AC	Q67JH4;			
DT	13-SEP-2005 (Rel. 48, Created)			
DT	13-SEP-2005 (Rel. 48, Last sequence update)			
DE	Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).			
GN	Name=hutH; OrderedLocusNames=STH3194;			
OS	Symbiobacterium thermophilum.			
OC	Bacteria; Actinobacteria; Symbiobacterium.			
OX	NCBI_TaxID=2734;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	STRAIN=T / IAM 14863;			
RX	PubMed=15383646; DOI=10.1093/nar/gkh830;			
RA	Ueda K., Yamashita A., Ishikawa J., Shimada M., Watsuji T.,			
RA	Morimura K., Ikeda H., Hattori M., Beppu T.;			
RT	"Genome sequence of Symbiobacterium thermophilum, an uncultivable bacterium that depends on microbial commensalism.";			
RL	Nucleic Acids Res. 32:4937-4944 (2004).			
CC	!- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).			
CC	!- PATHWAY: Histidine degradation; first step.			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).			
CC	!- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO), which is formed autocatalytically by cyclization and dehydration of residues Ala-Ser-Gly (By similarity).			
CC	!- SIMILARITY: Belongs to the PAL/histidase family.			
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.			
CC	EMBL; AP006840; BAD42176.1; -; Genomic_DNA.			
DR	HMAP; MF_00229; -; 1.			
DR	InterPro; IPR005921; HutH.			
DR	InterPro; IPR001106; Phe/His_NH3lyase.			
DR	InterPro; IPR005829; Sug_transporter.			
DR	Pfam; PF00221; PAL; 1.			
DR	TIGRFAMs; TIGR01225; hutH; 1.			
DR	PROSITE; PS00488; PAL_HISTIDASE; 1.			
KW	Complete proteome; Histidine metabolism; Lyase.			
FT	MOD_RES 143 143			
FT	CROSSLINK 142 144			
FT	SEQUENCE 507 AA; 53709 MW; 4343495574F09B5 CRC64;			
SQ	SEQUENCE 507 AA; 53709 MW; 4343495574F09B5 CRC64;			

Query Match 31.9%; Score 851; DB 1; Length 507;

Best Local Similarity 42.8%; Pred No. 2,1e-45;

Matches 217; Conservative 69; Mismatches 197; Indels 24; Gaps 9;


```
Qy 13 AVELDRHIDLDQAVANAGGARIIVLAPPARDRCRASEARLGAIVREARHVYGLTGTGPGPL 72
Db 3 AVELGAHLTLPEVVAHGARGVLLTPEVQRVARASEMVERLVRERRPVYGIITGFGKF 62

Qy 73 ANRLISGENVRLQANLVHHLASGVPLDWTTRAMVRLVLSIAQASGASGEGTARL 132
Db 63 SDVPISAEQTLQORNLMSHACAVGBPLAAEVVRAMLLLRQAQLSRHSIGIRAEITLML 122

Qy 133 IDLNSLAPAVPVRGSGDGLTPLAHMVLCLQGRGDFDRDGTTRLDGAEGRLRRGLQ 192
Db 123 VAFNLGLTPVVPQSGISGASDGLAPLAHMSLPLIGLGEAV-VNGERLSGAEALQVRGLR 181

Qy 193 PLDSLHRDALLVNGTSAMTGIALVNAHACHLGNWAVALTALLAECLRGRTAEWAALS 252
Db 182 PLTLTAKEGLALINGTOAMTALGSLGLHDAQVLLKTADIAAAMTAEALGAIPAAMDPRVQ 241

Qy 253 DLRPHPGKQAAARLARVDGSAVRVHVAERRLDAGDIGTEPEAGQDAYSLRCAPVL 312
Db 242 ALRUHTGQAAARNRLRITEGSRLLTTR-----PGQWRT-----ODPYTLRCLPQVH 287

Qy 313 GAGFDTTLAWHDRVLTIELNAVTDNP-VPPPDGSAVPLHGGNFMGQHVALTSDALATAVTV 371
Db 288 GASRTAIEHVAQVLDWENNAVTDNPLFPDDDEV--ISGGNFHGQPVVALDYLAIABAE 345

Qy 372 LAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRATG-PASIH 430
Db 346 LGDIAERRIERNLPOLS-GLPAFLTRN-GGVHSGMLTITVTAASLVSEKNVLAHPASVD 403

Qy 431 SISITNAQDVVSLGTIAARLCKREKIDRWAEILAILALCLAAQAEALRCGSLDGVSPAGK 490
Db 404 SIPSSAQEDHVSMGTAAARKARQVIANRVRVLAIELLCAQAQLEF---VGPERLAPATR 460

Qy 491 KLVALREQPPPLETRDPLGOEIAALA 517
Db 461 AYAARIRVAPLSGDRVLADPIEALA 487

RESULT 2
Q8GMG0 STRGL PRELIMINARY; PRT; 539 AA.
AC Q8GMG0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Putative ammonia lyase/transferase.
OS Streptomyces globisporus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C-1027;
RX MEDLINE=22171413; PubMed=12183628; DOI=10.1126/science.1072110;
RA Liu W., Christenson S.D., Standage S., Shen B.;
RT "Biosynthesis of the enediyne antitumor antibiotic C-1027.";
RL Science 297:1170-1173(2002).
DR EMBL; AY048670; AAL06680.1; -; Genomic_DNA.
DR HSSP; P21310; 1GKM.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0016211; F:ammonia lyase activity; IEA.
DR GO; GO:0004397; F:histidine ammonia-lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006548; P:histidine catabolism; IEA.
DR InterPro; IPR005921; huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Lyase; Transferase.
SQ SEQUENCE 539 AA; 58138 MW; E0CE3B75B579B1D9 CRC64;
```

```
Query Match 30.5%; Score 814.5; DB 2; Length 539;
Best Local Similarity 40.3%; Pred. No. 4.5e-43;
Matches 202; Conservative 73; Mismatches 201; Indels 25; Gaps 8;

Qy 34 RIVLAPPARDRCRASEAR--LGAVIREARHVYGLTGTGFGPLANRLISGENVRLQANLVH 91
Db 32 RATVDVPAESIAKAQKSREIFEGIAEQNIPIYGVTTGYGEMIMQVDKSEVELQTNLVR 91

Qy 92 HLASGVPVLDTWTTRAMVRLVLSIAQASGASEGTIARLIDLNSLAPAVPVRSGTVG 151
Db 92 SHAGVGPLPAEDARAIVARLNTLAKGHSVRPILERLAQVLYNRCITPAIPEISLG 151

Qy 152 ASGDLTPLAHMVLCLQGRGDFDRDGTTRLDGAEGRLRRGLQPLDSLHRDALLVNGTSAM 211
Db 152 ASGDLAPLSHVASTLIGEG-VYLRDGRPVETAQVLAERGIEPLELRFKEGLALINGTSGM 210

Qy 212 TGIALVNAHACHLGNWAVALTALLAECLRGRTAEWAALSDDL-RPHPGKQAAARLAR 270
Db 211 TGLGSLVVGRLAQAQAEIVTALLIEAVRGSTSPFLAEGHDIARPHGQIDTAANMRAL 270

Qy 271 VDGSAVRVHVAERRLDAGDIGTEPEAGQD-----AYSLRCAPQVLGAGFDTLA 320
Db 271 MRSGSLTVEHADLRRELQ-----KDEAGKDVORSEIYLOKAYSLRAIPQVVGAVRDTLY 325

Qy 321 WHDRVLTIELNAVTDNPVPPPDGSAVPLHGGNFMGQHVALTSDALATAVTVLGLAERQI 380
Db 326 HARHKLRIELNSANDNPLFPPEGKEI--FHGANFHGQPIAFAMDFVTITLTLQGLVLAERQI 383

Qy 381 ARUTDERLNRLGPPFLHRGPAGLNSGFMGAQVATALLAEWRATGPAISHISTNAANQD 440
Db 384 NRVNLRHLSYGLPEFLVSGDPLHSGFAGAQVATALVAENRTTIGPASTQVSPNSGNDQD 443

Qy 441 VVSLGTIAARLCKREKIDRWAEILAILALCLAAQAEALRCGSG-LDGVSPACKLVAQLREQ 499
Db 444 VVSNGLISARNARRVLSNNKILAVEYLAQAQAVDI---SGRFDGLSPAAKAYEAVRRL 500

Qy 500 FPELETDRPLGOEIAALATHL 520
Db 501 VPTLGVDRYMADDIELVADAL 521

RESULT 3
Q4NR19_9DEL1 PRELIMINARY; PRT; 508 AA.
AC Q4NR19;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Histidine ammonia-lyase (EC 4.3.1.3).
GN ORFNames=AdhDRAFT_1286;
OS Anaeromyxobacter dehalogenans 2CP-C.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Myxococcaceae; Anaeromyxobacter.
OX NCBI_TaxID=290397;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Anaeromyxobacter
RT dehalogenans 2CP-C.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2CP-C;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Anaeromyxobacter
RT dehalogenans 2CP-C.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
```



```
RESULT 5
ID HUTH GLOVI STANDARD; PRT; 514 AA.
AC Q7NCB3;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=glr3066;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacterales; Gloeobacter.
OX NCBI_TaxID=33072;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PCC 7421;
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Matsumoto A., Nakazaki N., Shimo S.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids.";
RL DNA Res. 10:137-145(2003).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -1- SIMILARITY: Belongs to the PAL/histidase family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; BA000045; GAC91007.1; -; Genomic DNA.
DR HSP; P21310; 1GKM.
DR HAMAP; MF 00229; -; 1.
DR InterPro; IPR005921; Huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Complete proteome; Histidine metabolism; Lyase.
FT MOD_RES 148 148 2,3-didehydroalanine (Ser) (By
FT similarity).
FT CROSSLINK 147 149 5-imidazolinone (Ala-Gly) (By
FT similarity).
FT SEQUENCE 514 AA; 54037 MW; 15CC63A8ED444DA4 CRC64;
Query Match 29.9%; Score 797.5; DB 1; Length 514;
Best Local Similarity 38.9%; Pred. No. 5e-42;
Matches 198; Conservative 80; Mismatches 202; Indels 29; Gaps 10;
QY 22 LQAHAVASGARTVLAPPARDRCSEARIGAVIREARHYGLTTGFGPLANRLTSGEN 81
DB 17 VDLIVAVARGVPVPSASLELVRRSRFAVEALLEGDEIVGTTGFGFKRRRPRSA 76
QY 82 VRTLQANLVHHLASGVGVLDWTTARAVLARLVSTIAQAGSAGSEGTIARLIDLNSLA 141
DB 77 VEQLQQLNMSSAAGLGPFGREVVRAMLLLRANTLAQGYSGVRPETLQLLVAMLRGVH 136
QY 142 PAVPSRGTVCASGDLTPLAHMVLCLQGRGDFLRDGRDLGAGRLGRLOPLDLSHRDA 201
DB 137 PVVPCRGSVGSDGLAPLAHLALVLTGEGE-AEYGVGEVLPFAAALARAGLEPIRLGAKG 195
QY 202 LALVNGTSMTGTALVNAHACRHGNWAVALTALAECLRGRTAEWAASLDRPHPGQK 261
DB 196 LALINGTQMSALGALTVTHRAQRLAKLADLACANTLEATIGRSFAFLPHFRLRPHGQQ 255
QY 262 DAAARLRARVDGSARVVRVHIAERRLDAGDIGTEPEAGQDAYSILRCAPQVLGAGFDTLAW 321
```

```
Db 256 SSARNLLVLTEDSALITASHAGCDR-----VQDAYSLRCAPQVHGASLDAISY 302
QY 322 HDRVLTIELNAVTDNP-VPPDGSVPALHGGNFMGQHVLTSDALATAVTVTLAGLAERQI 380
Db 303 AAGVIAIEINSDNPLIFADTQGV--VTGGHFGQPVAVASDVLAIALAELADISERRT 360
QY 381 ARLTDERLNRGLPPFLHRGPGAGLNSGFMGAQVTTATALLAEMRATG-PASIHSTSTNAANQ 439
Db 361 ERLVNADYSNGLPMFLTEA--GLHSGYVVAQYTAASLVSENVKVLAHFACVDSIPTSAQGE 419
QY 440 DVVSLGTIATARLCKEKIDRWAEILAILALCIAQAELURCGSLDGVSPAGKUL-VQALRE 498
Db 420 DHVSMGLTARKAVTVCDCNCEVIAELMCAQAQALDLR-GKLTTPG---RGSRVGLEVIRA 475
QY 499 QFPLETRDPLGOEIAALAT-----HLIQ 522
Db 476 AVPHLSSDRIVSRDIEKVVLMADGHLLE 504
RESULT 6
Q9KHJ9_9ACTO
ID Q9KHJ9_9ACTO PRELIMINARY; PRT; 523 AA.
AC Q9KHJ9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Putative phenylalanine ammonia lyase Encp.
DE Names=Encp;
GN Streptomyces maritimus.
OS Streptomyces maritimus; Actinobacteridae; Actinomycetales;
OC Bacteria; Actinobacteria; Actinobacteridae; Streptomyces.
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=115828;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Piel J., Hoang K., Moore B.S.;
RT "Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis
RT Gene Cluster.";
RL J. Am. Chem. Soc. 122:5415-5416(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20579517; PubMed=11137817; DOI=10.1016/S1074-5521(00)00044-2;
RA Piel J., Hertweck C., Shipley P.R., Hunt D.M., Newman M.S.,
RA Moore B.S.;
RT "Cloning, sequencing and analysis of the enterocin biosynthesis gene
RT cluster from the marine isolate 'Streptomyces maritimus': evidence for
RT the derailment of an aromatic polyketide synthase.";
RL Chem. Biol. 7:943-955(2000).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21942282; PubMed=11948863;
RX DOI=10.1002/1439-7633(20011001)2:10<784::AID-CBIC784>3.0.CO;2-K;
RA Hertweck C., Jarvis A.P., Xiang L., Moore B.S., Oldham N.J.;
RT "A mechanism of benzoic acid biosynthesis in plants and bacteria that
RT mirrors fatty acid beta-oxidation.";
RL Chembiochem 2:784-786(2001).
DR EMBL; AF254925; AAF81735.1; -; Genomic DNA.
DR HSP; P21310; 1GKM.
DR GO; GO:0016211; F:ammonia lyase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
KW Lyase.
SQ SEQUENCE 523 AA; 56416 MW; 012440F937D48FD9 CRC64;
Query Match 29.7%; Score 792.5; DB 2; Length 523;
Best Local Similarity 35.7%; Pred. No. 1.1e-41;
Matches 186; Conservative 101; Mismatches 217; Indels 17; Gaps 8;
QY 14 VELDRHIDLDOAHAVASGGARIVLAPARDRCSEARIGAVIREARHYGLTGGFLA 73
DB 5 IELDMNVTLQLEBDAARQRTFVELSAPVRSRVRASRDVLVKFVODERVIYGVNTSMGGFV 64
```

QY 74 NLLISENVRTLOANLVHLLASGVPLDWTTRAMVRLARLVSIQAQASGASEGTIARLI 133
 Db 65 DHLVPSVQARQOENLINAATVNGAYLDDTTARTIMLSRIVSLARGNSAITPANLDDKLV 124
 QY 134 DLNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGLDCAEGLRGRLOP 193
 Db 125 AVLNAGIIVPCIEKSGSLTSGDLPLAALAVCAQWK-ARYNGQIMPGRQALSEAGVEP 183
 QY 194 LLSHRDALALVNGTSAMTGIALVNAHACRHGNVAVALTALLAECLRGRTTEMAAALS 253
 Db 184 MELSXYKDGGLALNGTSGMVGLTVMVLAARLVDRYLOVSALSVEGLAGMTKPPDPVRHG 243
 QY 254 LRPHPGQDAARLARARVDGSARVVRHVIAERLDAGDGTGTEPEAG-----QDAYSLRCAP 309
 Db 244 VAPHRGQVARSRLMEGLADSHLANVELDTEQLT-AGEMGTAKAGSLAIEDAYSIRCTP 302
 QY 310 QVLGAGFDTLAWHDRV---LTIELNAVTDNVPFPPDGSVPALHGGNFMGQHVALTSDALA 366
 Db 303 QILGPVVDVL---DRIGATLQDELSSNDNPVLPE-EAEVPHNGHFGQVYAMAMDHNL 358
 QY 367 TAVTVLAGLAERQIARLTDERLNRLGPPFLHRRGAGLNSGFMGAQVTTATALLAEVRA-TG 425
 Db 359 MALATVTNLANRRVDRFLDKSNSNGLPAFLCREDPGLRLGLMGQGFMTASITAEITRTITI 418
 QY 426 PASIHSISNAANDVSLGTIAARLCREKIDRWAEILAILALCLAAELRCGSGLDGV 485
 Db 419 PMSVQSLSTADFODIVSFGFVAARRAREVLNTAAAYVAVFELLCAQAVDIR---GADKL 475
 QY 486 SPAGKKLVQALREQPPPLETRPLGQETAAALATHLLQOSP 526
 Db 476 SSFTRPLRYERTKIVPFDRDETITDYVEKLAADLIAGEPV 516

RESULT 7

Q4H6E7 9DEIO
 ID Q4H6E7 9DEIO PRELIMINARY; PRT; 499 AA.
 AC Q4H6E7
 DT 13-SEP-2005 (TRENBLrel. 31, Created)
 DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
 DE Histidine ammonia-lyase (EC 4.3.1.3).
 GN ORFNames=DgeODRAFT_2894;
 OS Deinococcus geothermalis DSM 11300.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Rhodocyclaceae; Deinococcus.
 OX NCBI_TaxID=319795;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=DSM 11300;
 RG US DOE Joint Genome Institute (JGI-PGF);
 RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
 RA Hammon N., Israni S., Pittluck S., Richardson P.;
 RT "Sequencing of the draft genome assembly of Deinococcus geothermalis
 DSM 11300.";
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=DSM 11300;
 RG US DOE Joint Genome Institute (JGI-ORNL);
 RA Larimer F., Land M.;
 RT "Annotation of the draft genome assembly of Deinococcus geothermalis
 DSM 11300.";
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
 CC -! CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAHE01000014; EAL61898.1; -; Genomic_DNA.
 KW Lyase.
 SQ SEQUENCE 499 AA; 52368 MW; CA03423FDA15AC62 CRC64;

Query Match 28.6%; Score 763; DB 2; Length 499;
 Best Local Similarity 40.0%; Pred. No. 7.2e-40;

Matches 200; Conservative 69; Mismatches 195; Indels 36; Gaps 12;
 QY 16 LDRHIDLQAHAVASGGGARIVLAPPARDRCRASEARLGAIVIREARHVYGLTGTGFGPLANR 75
 Db 3 LDRQLTLDDFIRVRVGGEVTLDAARTRMGARAVIERI VDGPEAVYGVNTGFGKFAV 62
 QY 76 LISGENVRTLQANLVHLLASGVPLDWTTRAMVRLARLVSIQAQASGASEGTIARLI 135
 Db 63 RVAREBELQOLHNLIVSHAIGVAGLPAEVVRGMULLRAQSLALGHSGVVRPEVVELLLAL 122
 QY 136 LNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGLDCAEGLRGRLOP 195
 Db 123 LNAGACFPVPAQSGVSGDLPLAHLALALIGGGE-LEYGQVPRPADVLAEGLQPLIT 181
 QY 196 LSHRDALALVNGTSAM---TGIALVNAHACRHGNVAVALTALLAECLRGRTTEMAAALS 252
 Db 182 LEAKEGLALINGTQLMGSLALLALHDARTLLHTANLAAMT---VEALSGSHRPFSEGVV 238
 QY 253 LDRPHPGQDAARLARARVDGSARVVRHVIAERLDAGDGTGTEPEAGODAYSILRCAPVL 312
 Db 239 SLRPHPGALEVAADLRFLHGS-----IAPAHACGKV-----QDAYSLRAVPOVH 285
 QY 313 GAGFDTLAWHDRVLTIELNAVTDN-VEPPDGSVPALHGGNFMGQHVALTSDALATATVTV 371
 Db 286 GASLDALMQAGRVLEVEFASVTDNPLIFPETGEV--ISGNEFHGQPLAALADALKVAVAE 343
 QY 372 LAGLAERQIARLTDERLNRLGPPFLHRRGAGLNSGFMGAQVTTATALLAEVRA-TG-PASTH 430
 Db 344 LANISERSEQLLNPLALS-GLPGFL-TPEGGLSSGFMIAQYTAALVSENKVLHAPASVD 401
 QY 431 SISITNAANDVSLGTIAARLCREKIDRWAEILAILALCLAAELRCGSGLDGVSP 487
 Db 402 SIPTSANQEDHVSMAHGARGQLRQILENAQSVISIELLCAAQALDFQSLRAGRGVQAA--- 459
 QY 488 AGKKLVQALREQPPPLETRD 507
 Db 460 -----YERIRQEVAPLQDQR 474
 ID HUTH AZOSE STANDARD; PRT; 526 AA.
 AC QSNZXB;
 DT 13-SEP-2005 (Rel. 48, Created)
 DT 13-SEP-2005 (Rel. 48, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
 GN Name=huth; OrderedLocusNames=AZOSEA32610; ORFNames=eBA5742;
 OS Azococcus sp. (strain EDN1).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales;
 OC Rhodocyclaceae; Azococcus.
 OX NCBI_TaxID=761114;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RA PubMed=15551059; DOI=10.1007/s00203-004-0742-9;
 RA Rabus R., Kube M., Heider J., Beck A., Heitmann K., Widdel F.,
 RA Reinhardt R.;
 RT "The genome sequence of an anaerobic aromatic-degrading denitrifying
 bacterium, strain EDN1.";
 RL Arch. Microbiol. 183:27-36 (2005).
 CC -! CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
 CC -! PATHWAY: Histidine degradation; first step.
 CC -! SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -! PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
 CC which is formed autocatalytically by cyclization and dehydration
 CC of residues Ala-Ser-Gly (By similarity).
 CC -! SIMILARITY: Belongs to the PAL/histidase family.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.

```
CC EMBL; CR555306; CAI09386.1; -; Genomic_DNA.
DR HAMAP; MF_00229; -; 1.
DR InterPro; IPR005921; huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Complete proteome; Histidine metabolism; Lyase.
FT MOD_RES 144 144
FT CROSSLINK 143 145
FT 5-imidazolinone (Ala-Gly) (By
FT similarity).
FT SEQUENCE 526 AA; 54655 MW; 94B1E777DC1574D3 CRC64;
Query Match 28.6%; Score 762.5; DB 1; Length 526;
Best Local Similarity 40.1%; Pred. No. 8.3e-40;
Matches 200; Conservative 65; Mismatches 211; Indels 23; Gaps 9;
Qy 20 IDLQAHAVASGGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGPLANRLISG 79
Db 11 LTLAELRTIAPSDSRLELEPACFPVARGAATVAIARSGEPAYGINTGFGRLAQTHIPD 70
Qy 80 ENVTTLQANLVHHLASGVGVLDWTTTARAMVLARLVIAQASGSGTTIARLIDLNSE 139
Db 71 DQELLEQNLVLSHAVGVGEPLSAPTIVRLVLKIALKSARGHSGVRMELINALIGLFNAG 130
Qy 140 LAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRTGDLDAEGLRRCGRLOPLDLSHR 199
Db 131 VIPRPSKSGVSGASGDLAPLAHLASALLGIGIYAY-VDRGHVPATEALAIAGLAPMTLAAK 189
Qy 200 DALALVNGTSMNTGIALVNAHACHRLGNVAVALTALLAECLRGRTAEWAALSLDRPHG 259
Db 190 EGLALLNGTQVSTALALVNLFAIETVFTALVAGALSVDAAAGSFKFPDARIHALRQPG 249
Qy 260 QXDAARLARVDSGARVVRHVIAERLDAGDIGTEPEAGQDAYSRLCARQVTLGAGFDTL 319
Db 250 QIDAAATYRQLLEGSGINLAH-----RDCGKV-----QDPYSRLCQPVVMGACLDQM 296
Qy 320 AWHDRVLTIENAVTDNP-VPPDGSVPALHGGNFMGQHVALTSDALATATVTLAGLAER 378
Db 297 RHAARVLLIEANAVSDNPLVPFDSGEV--LSGGNFHGEVPVAFADALALAAAEIGALAE 354
Qy 379 QIARLTDERLNRLGPPFLHRGPAGLNGFGMGAQVATATALLAEMRATG-PASIHISISTNAA 437
Db 355 RIALLIIDATLS-GLPPFLVT-EGVNSGFMIAHVTAALASSENKLLAHFASVDSLPFSAN 412
Qy 438 NQDVVSLGTIARLCREKIDRWAEILAILALCLAAQAAELRCGSLDGVSPAGKKLVQALR 497
Db 413 QEDHVSMTFAARKLGELADNTATILAILLAAAGVVELRAP---HRTSPRLQAVLALIR 469
Qy 498 EQPPLTDRPLGQEIALL 516
Db 470 SRVPHYDIDRYFAPDIASI 488
```

RESULT 9

```
ID HUTH_STIAU STANDARD; PRT; 510 AA.
AC Q93TX3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth;
OS Stigmatella aurantiaca.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cystobacterineae; Cystobacteraceae; Stigmatella.
OX NCBI_TaxID=41;
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RX MEDLINE=2110452; PubMed=11182319; DOI=10.1016/S1074-5521(00)00056-9;
RA Silakowski B., Nordsiek G., Kunze B., Blocker H., Muller R.;
```

```
RT "Novel features in a combined polyketide synthase/non-ribosomal
RT peptidase synthetase: the myxalamid biosynthetic gene cluster of the
RL myxobacterium Stigmatella aurantiaca Sga15."
CC Chem. Biol. 8:59-69(2001).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; AF319998; AAK57183.1; -; Genomic_DNA.
DR HESP; P21310; IGKM.
DR HAMAP; MF_00229; -; 1.
DR InterPro; IPR005921; huth.
DR InterPro; IPR001106; Phe/His_NH3lyase.
DR Pfam; PF00221; PAL; 1.
DR TIGRFAMs; TIGR01225; huth; 1.
DR PROSITE; PS00488; PAL_HISTIDASE; 1.
KW Histidine metabolism; Lyase.
FT MOD_RES 146 146
FT 2,3-didehydroalanine (Ser) (By
FT similarity).
FT CROSSLINK 145 147
FT 5-imidazolinone (Ala-Gly) (By
FT similarity).
FT SEQUENCE 510 AA; 54201 MW; DE78439DB4FBF3A6 CRC64;
Query Match 28.2%; Score 753; DB 1; Length 510;
Best Local Similarity 38.3%; Pred. No. 3.2e-39;
Matches 198; Conservative 85; Mismatches 202; Indels 32; Gaps 10;
Qy 7 MSPKPAVALDRHIDLDQAHAVASGGARIVLAPPARDRCRASEARLGAVIREARHVYGLT 66
Db 1 MSFRLNIDGDT-LKLEELQVARTVTVELAPAAARVCAARDLVDRVAAGDTPSGIN 59
Qy 67 TGRGPLANRLISGENVTLQANLVHHLASGVGVLDWTTTARAMVLARLVIAQASGASE 126
Db 60 TGFGLTAEVRIDKDLRELQNLILSHAAGVSGPLPLPEARVLLLRNCNVLAKGYSGRP 119
Qy 127 GTTARLIDLNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRTGDLDAEG 186
Db 120 ETLTALALEMLNRDVPVWPERGSGASGDLAPLAHLALVFIEGEAFYK-GERLPAAQAL 178
Qy 187 RRCGRLOPLDLSHRDALVNGTSMNTGIALVNAHACHRLGNVAVALTALLAECLRGRTAE 246
Db 179 ERAGLKPVLLEAKEGLALVNGTQMCVGTILQLRAEMLADLADAGMTLEGLLSHKP 238
Qy 247 WAAALSDLRPHPGQKDAARLARVRDGSARVVRHVIAERLDAGDIGTEPEAGQDAYSRLR 306
Db 239 FIPEIQDVRAHEGQKACAAHRELLADLSALVESHVNCCK-----VQDPYSRLR 285
Qy 307 CAPQVLGAGFDTLAWHDRVLTIELNAVTDNP-VPPDGSVPALHGGNFMGQHVALTSDAL 365
Db 286 CMPQVHGAAREGLSPARRILEVEINSATDNPVLFVETERI--VSGGNFHGQPVSLADVA 343
Qy 366 ATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNGFGMGAQVATATALLAEMRA-T 424
Db 344 AMALTQLSAISERRVQLVNPALS-GLPPPLAKN-SGLNSGFMIAQVTSALVAESRVL 401
Qy 425 GPASIHISITNAANQDVSLGTIARLCREKIDRWAEILAILALCLAAQAAE-----LRCGS 480
Db 402 HPASVDSIPSAGREDHVSNGMTAALKGRQVADFTRSCLAIELLVAQAALDYQPTPRAK 461
Qy 481 GLDGVSPAGKKLVQALREQPPPLETRPLGQEIALLA 517
Db 462 G-----PQAAEYLIRSKIPTMEKORELHRDIAAVS 491
```

```

RESULT 10
HUTH_BURMA STANDARD; PRT; 507 AA.
ID AC Q621J6;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocNames=BMA0645;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Niernan W.C., Deshazer D., Kim H.S., Tettelin H., Nelson K.B.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M.,
RA Daugherty S.C., Davidson I.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -1- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; CP000010; AAU49673.1; ALT_INIT; Genomic_DNA.
CC TIGR; BMA0645; -.
CC HAMAP; MF_00229; -.
CC InterPro; IPR005921; Huth.
CC InterPro; IPR001106; Phe/His_NH3lyase.
CC Pfam; PF00221; PAL; 1.
CC TIGRFAMs; TIGR01225; huth; 1.
CC PROSITE; PS00488; PAL_HISTIDASE; 1.
CC Complete proteome; Histidine metabolism; Lyase.
CC MOD_RES 142 142 2,3-didehydroalanine (Ser) (By
CC similarity).
CC CROSSLINK 141 143 5-imidazolinone (Ala-Gly) (By
CC similarity).
CC SEQUENCE 507 AA; 53141 MW; 77B4E49060B6D4AE CRC64;
Query Match 28.1%; Score 749.5; DB 1; Length 507;
Best Local Similarity 38.4%; Pred. No. 5.2e-39;
Matches 202; Conservative 74; Mismatches 213; Indels 37; Gaps 11;
QY 4 MLAMSPKPAVELDRHLDLOAHAVAGGARIVLAP---PARDRCASEARLGNVREAR 60
DB 1 MITLTPGR-----LTLPLQRIARENVQIALDPASFAAIDRAQAQADIAA---KGE 49
QY 61 HVYGLTGFGLANRLISGENVRTLOANVHLHLSAGVGPVLDWTTARAVLARLVSTAQ 120
DB 50 PAYGINTGFGRLASTHPIHQLELLQKNLVLSHAGVGSEMPARVPRLLMALKLSSLG 109
QY 121 ASGASECTIARLIDLNLSEAPAVPSRGTVGASGDLTFLAHMVLCLQGRGDFLDRDCTRL 180
DB 110 HSGIRRVMDALVTLFNADVLPLIPVKGSGASGDLAPLAHMSAVLIGDVFIR-GERA 168
QY 181 DGAEGLRGRQLQDLDRDALALVNGTSAMTGTIALVNAHACRHGLGNVAVALTALLAECL 240

```

```

169 SAABGLRVAGLAPLTLEAKGGLALINGTQASTALADNLFAIEDLYRTALVSGALSVDA 228
241 RGRTEAWAAALSDLRPHGPQKDAARLARVRVDGSRVVRHVIAERRLDAGDITGEPEAQ 300
229 AGSVKPPDARITHELGRHGQIDAAAAYSLDGSAINVSH-----RCDKQV 275
301 DAYSIRCAPOVLGAGFTFLAWHDRVLTIELNAVTDNP-VFPDGSVPALHGNFQHWVA 359
276 DPLSLRCOPQVMGACLDQIRHAAGVLLTEANAVSDNPLIFPDPTGEV--LSGNGFHAEP 333
360 LTSALATATVTLAGLAEQIARLTDEBLNGLPFLHGPAGLNSGFWGQVTTATALLA 419
334 PAADNLATAAAEIGALAEIRIALLIDATLS-GLPFLVK-DGVNNGFMIAHVTTAAALAS 391
420 EMRATG-PASIHSTISNAANQDVVSLGTIAARLCREKIDRWAEIILAILALCIAAEALRC 478
392 ENKTLAPASVDSLPTSANQEDHVSMTAFARKLTDAENVANILAIELLAAGQVDLRA 451
479 GSGLDGVSPPAGKLVQALREQFPLETDRLPGQIBIATLHLLQOS 524
452 P---HATSPALQHAMKTIRADVADHYLDHYPADPAIVAVRVRERA 494
RESULT 11
HUTH_BURPS STANDARD; PRT; 507 AA.
AC Q63SH6;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocNames=BPSL2344;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdeno-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., DeShazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songsvilai S., Stevens K., Tumapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (by similarity).
CC -1- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; BX571965; CAH36347.1; -; Genomic_DNA.
CC HAMAP; MF_00229; -.
CC InterPro; IPR005921; Huth.
CC InterPro; IPR001106; Phe/His_NH3lyase.
CC Pfam; PF00221; PAL; 1.

```



```
Db 172 DGLRVAGLAPUTLQAKGLALNGTQASTALDNLMEFAIEDLYRTALVAGALSVDAAAGS 231
Qy 244 TEAWAALSDLRPHPGKDAARLARVDGSRVVRHVIARRRDLADGIGTEPEAGQDAY 303
Db 232 VKPFDARIELRGHQGIDAASVRELLEGSF-----INQSHRDCVK-----QDPY 278
Qy 304 SLRCAPQVLGAGFTPLAWHDRVLTIELNAVTDNP-VFPDGSVPALHGGNFMGQHVALT 362
Db 279 SLRCOPQVMGACLDQMRHAADVLLVEANVSDNPLIFPDTEGV--LSSGNFHAEPVAFAA 336
Qy 363 DALATATVVLGAEARQIARLTDELNRGLPPFLHRGPAGLNSFGMCAQVTTALLAEMR 422
Db 337 DNLALAAEIGALAEIRIALIDATLS-GLPFPFLVK-DGGVNSGFMFAHVTAAALASENK 394
Qy 423 ATG-PASHSISTWAQNVDSLGTIAARLCREKIDRWAEITLALCLQAELRCGSG 481
Db 395 TLAHPASVDSLUFTSANGEDHVSMTAFKRLADTADNTKHTLLELAAAGGVDLRAPY- 453
Qy 482 LDGVSFAGKLVQALRQFPFPLETRPLGQBIAAALA 517
Db 454 --HTSPKLPWMEITRGKVAHYELDHVFPADIAVIA 487

RESULT 15
HUTH_BACSU
AC P10944: STANDARD; PRT: 508 AA.
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUN-1989 (Rel. 11, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN Name=huth; OrderedLocusNames=BSU39350;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=88257040; PubMed=2454913;
RA Oda M., Sugishita A., Furukawa K.;
RT "Cloning and nucleotide sequences of histidase and regulatory genes in
RT the Bacillus subtilis hut operon and positive regulation of the
RT operon.";
RL J. Bacteriol. 170:3199-3205 (1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=168 / BGSC141;
RX MEDLINE=95219089; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapA loci.";
RL Microbiology 141:337-343 (1995).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Besieres P., Bolotin A., Borchert S.,
RA Borriis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.-K., Codani J.-J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Duesterhoeft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haich J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.-F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Iardinois S., Lauber J.,
RA Lazarevic V., Lee S.-M., Levine A., Liu H., Masuda S., Mauel C.,
RA Medigue C., Meene N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
```

```
RA Park S.-H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone P., Sekiguchi J., Sekowska A., Seror S.J., Serron P.,
RA Shin B.-S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takanashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzneger T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.-F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
CC -I- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -I- PATHWAY: Histidine degradation; first step.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -I- INDUCTION: By histidine.
CC -I- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -I- SIMILARITY: Belongs to the PAL/histidase family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; M20659; AAA22538.1; -; Genomic DNA.
CC EMBL; D31856; BAA06644.1; -; Genomic DNA.
CC EMBL; Z99124; CAB15971.1; -; Genomic DNA.
CC PIR; S18810; UFBESH.
CC HSP; P21310; IGKM.
CC Subtilist; EG10667; huth.
CC HAMAP; MF_00229; -; 1.
CC InterPro; IPR005921; Huth.
CC InterPro; IPR001106; Phe/His_NH3lyase.
CC Pfam; PF00221; PAL; 1.
CC TIGRFAMS; TIGR01225; huth; 1.
CC PROSITE; PS00488; PAL_HISTIDASE; 1.
CC Complete proteome; Histidine metabolism; Lyase.
FT MOD_RES 142 142 2,3-dihydroalalanine (Ser) (By
FT CROSSLINK 141 143 5-imidazolinone (Ala-Gly) (By
FT FT similarity).
SQ SEQUENCE 508 AA; 55675 MW; 869C323BFCC318E0 CRC64;
Query Match 27.3%; Score 728.5; DB 1; Length 508;
Best Local Similarity 35.5%; Pred. No. 1.1e-37;
Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9;
Qy 20 IDLDQAHAVASGGARIVL-----APPARDRCRASEARLGAIVREARHYVGLTGTG 72
Db 2 VTLDGSSLLTADVARVLDFEFAAASESMERVKSRAAVERIVRDEKTIYGTGKGF 61
Qy 73 ANRLISENVRTLOANLVHHLASGVPLDWTARTAMVLARLVSIAGQASGEGTIARL 132
Db 62 SDVLIQKEDSAALQLNLILSHACGVDPFPECYSRAMLLRLRANALLKGFSGVRAELIEQL 121
Qy 133 IDLNLSELAPVPSRGTGVSGBLTPLAHMVLCLQGRGDFLDRDGLDGAELRGRLQ 192
Db 122 LAFENKRVFPIVQQSGSLGSGDLAPLSHLALALIQGGEVP-FEGEMPMANTGLKKAGIQ 180
Qy 193 PLDSLHRDALVNGTSAMTGIALVNAHACRHIGNNVAVALTALLAECLRGRTFAWAALS 252
Db 181 PVTLTSEGLALINGTOAMGVVYIAEKLAYQTERIASLTIEGLQGIIDAFEDIH 240
Qy 253 DLRPHFGQKDAARLARVDGSRVVRHVIARRRDLADGIGTEPEAGQDAYSLRCAPQVL 312
Db 241 LARGYEQIDVAERIRFYLSDSLTTTSQ--GELRV-----QDAYSLRCIPQVH 286
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 08:57:25 ; Search time 46 Seconds
(without alignments)
945.378 Million cell updates/sec

Title: US-10-621-826-3
Perfect score: 2668
Sequence: 1 VKPMLAMSPKPAVELDRHI.....RPLQEIATLATHLLQOSPV 526

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgm2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgm2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgm2_6/ptodata/1/iaa/PCUTS_COMB.pep.*
5: /cgm2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774.5	29.0	513	2	US-09-833-745-49
2	761	28.5	508	2	US-09-902-540-15192
3	749	28.1	513	2	US-09-833-745-52
4	749	28.1	513	2	US-09-833-745-58
5	728.5	27.3	508	2	US-09-833-745-40
6	724	27.1	513	2	US-09-833-745-46
7	721.5	27.0	513	2	US-09-833-745-44
8	717.5	26.9	513	2	US-09-833-745-50
9	710.5	26.6	513	2	US-09-833-745-56
10	706	26.5	504	2	US-09-833-745-65
11	706	26.5	516	2	US-09-833-745-41
12	695.5	26.1	678	2	US-09-949-016-10001
13	694.5	26.0	657	2	US-09-949-016-6174
14	690	25.9	513	2	US-09-833-745-45
15	687	25.7	513	2	US-09-833-745-41
16	687	25.7	525	2	US-09-540-236-2855
17	684.5	25.7	513	2	US-09-833-745-47
18	684.5	25.7	513	2	US-09-833-745-48
19	683	25.6	508	2	US-09-489-039A-13575
20	681	25.5	513	2	US-09-833-745-53
21	680.5	25.5	513	2	US-09-833-745-61
22	678.5	25.4	513	2	US-09-833-745-57
23	676.5	25.4	513	2	US-09-833-745-59
24	676.5	25.4	513	2	US-09-833-745-60
25	674	25.3	513	2	US-09-833-745-55
26	672.5	25.2	635	2	US-09-833-745-38
27	670	25.1	677	2	US-09-833-745-39

28	666.5	25.0	657	2	US-09-833-745-37	Sequence 37, Appl
29	662.5	24.8	657	2	US-09-833-745-36	Sequence 36, Appl
30	660.5	24.8	502	2	US-09-833-745-66	Sequence 66, Appl
31	660.5	24.8	513	2	US-09-833-745-5	Sequence 5, Appl
32	660.5	24.8	513	2	US-09-833-745-43	Sequence 43, Appl
33	654	24.5	515	2	US-09-328-352-5090	Sequence 5090, Ap
34	650	24.4	513	2	US-09-833-745-62	Sequence 62, Appl
35	647	24.3	513	2	US-09-833-745-51	Sequence 51, Appl
36	643	24.1	509	2	US-09-833-745-34	Sequence 34, Appl
37	631.5	23.7	511	2	US-09-833-745-10	Sequence 10, Appl
38	631.5	23.7	511	2	US-09-833-745-42	Sequence 42, Appl
39	626	23.5	511	2	US-09-833-745-35	Sequence 35, Appl
40	624	23.4	515	2	US-09-252-991A-33046	Sequence 33046, A
41	617	23.1	513	2	US-09-833-745-54	Sequence 54, Appl
42	616.5	23.1	520	2	US-09-252-991A-33049	Sequence 33049, A
43	584	21.9	513	2	US-09-833-745-63	Sequence 63, Appl
44	573	21.5	405	2	US-09-833-745-4	Sequence 4, Appl
45	559.5	21.0	404	2	US-09-833-745-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-833-745-49
; Sequence 49, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus subtilis
; US-09-833-745-49

Query Match	29.0%	Score	774.5	DB 2	Length	513			
Best Local Similarity	36.4%	Pred. No.	3.5e-67						
Matches	187	Conservative	94	Mismatches	206	Indels	27	Gaps	7
QY	20	IDLQAHAVASGARIVL-----APPARDRCRASEARLGAVIREARHVYLTGTGQPL	72						
DB	7	VTLDGSSLTTADVARVLDFEEAAASESMERVKSRAAVERIVRDEKTIYINGTGFQK	66						
QY	73	ANRLISGENTVTLQANLVHHLASGVPLDWTTRAMVLRVLSIAQASGASEGTIARL	132						
DB	67	SDVLQKEDSAALQANLILSHACGVGDPFPECVSRAMLLLRANALLKGFSGVRAELTEQL	126						
QY	133	IDLLNSLAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTLDGAEGLRRGLQ	192						
DB	127	LAFENKRVHPVI PQQSLGASGDLAPLSHLALALIGQEVDFEGERMPAMTGLKKAGIQ	186						
QY	193	PLDLSHRDALALVNGTSAMTGIALVNAHAACHLGNNAVALTALLAECILRGRTAEWAALS	252						
DB	187	PVTLTSKEGALINGTQAMTAMGVWAYIEAKLAYQFERTIASLTIEGLQGIIDAFDEDIH	246						
QY	253	-DURPFGQKDAARLARVDGSGARVVRVIAERRLDAGDIGTEPEAGQDAYSRCAPQV	311						
DB	247	LALRGYQEQIDVAERIRFYLSDSLTSRQGLRV-----QDAYSRLRCIPQV	294						
QY	312	LGAGFDTLAWHDRVLTIELNATVDNVPFPPDGSVPALHGGNFMGSHVALTSDALATAVTV	371						
DB	295	HGATWTQTLGYVKLEIEMNAATDNPLIFNDGDV--ISGGNFGHQPITAFAMDFLKIAISE	352						

```
QY 372 LAGLAERQIARLTDERLNRGLPPLHRRGAGLNSGFMGAQVTTALAEARATG-PASIH 430
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 353 LANIAERIERLNPQLNRDLPPFLSPHP-GLQSGAMIMQYAAASUSVSENKTLAHPASVD 411
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 431 SISTNAANQDVSVLGTITAAKLCREKIDRWAEIILAILCALQAABEUCGSLDGVSPAGK 490
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 412 SIPSSANQEDHVSNGTIAARHAYQVIANTRRVIAIEALCALQAVEVR--GIEHAASYTK 468
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 491 KLVAQALREQPPLETRDPLGQETAAALATHLQOS 524
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 469 QLFQEMKRVPSIQODRVFSYDIERLTDWLKGS 502
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 2
US-09-502-540-15192
; Sequence 15192, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Goldstein, Gregory J.
; APPLICANT: Hinkle, Steven C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15192
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-502-540-15192

Query Match 28.5%; Score 761; DB 2; Length 508;
Best Local Similarity 38.8%; Pred. No. 7.3e-66;
Matches 200; Conservative 84; Mismatches 202; Indels 30; Gaps 10;

QY 7 MSPPKPAVELDRHLDQAHAVASGGARIVLAPPARDRCRASEARLCAVIREARHVGLT 66
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSRPRLIDGDT-LKLEIILQVARNETVLSFDAATRVASRALVDRVAAGTTPAYGIN 59
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 67 TGFQPLANRLISGENVRLTQANLVHLSASGVPLDWTARTAMVLAARLVLSIAQASGASE 126
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TGFQTLAEVRIDKDLQDLQNLILSHACGVGTPLPLPEARALLLCNVLAAGYSGIRM 119
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 127 GTIARLIDLNSBELAPVPSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGTLDGAEGL 186
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 ETALALDMLNRDVPVPPERGSGVSGDGLAPLAHLALVFIGEAP-YQQQMPAKQAL 178
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 187 RRGRLQPLDLSHARDALANGTSAMTGIALVNAHACHRLGNWAVALTALLAECLRGTEA 246
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 ERAGLPQVLEAKLEGALVNGTQAMCAVGTLLQRAESLADIADVAGAMTEGLLGSHKP 238
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 247 WAAALSDLRPHQKQAAALRLARVDGSARVRVHVAERLADGDTGTEPEAQDAYSRLR 306
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 FIPEIHDRVRAHPGQKQVAALRLILVDSELVSHVNCSE-----VQDPYSUR 285
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 307 CAPQVLGAGPDTLAWHDRVLTIELNAVTONPVPDPGSPALHGGNFMGCHVALTSDALA 366
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 286 CMPQVHGAREGIAFSEIRILEVEVNSATDNPVAFAD-TERIVSGGNFHGQPISLAMDVVA 344
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 367 TAVTVLAGLAERQIARLTDERLNRGLPPLHRRGAGLNSGFMGAQVTTALAEARATG 425
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 345 MALTQLSISIRERVEQVNPVLSL-NLPAFLAKN-SGLNSGFMIAQVTSAAALVAESVLSH 402
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 426 PASIHSITSTNAANQDVSVLGTITAAKLCREKIDRWAEIILAILCALQAABEUCR- 481
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 403 PASVDSIPSSAGREDHVSNGTIAARHAYQVIANTRRVIAIEALCALQAVEVR--GIEHA 462
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 482 LDGVSPAGKKLVQALREQFPPLPPLHRRGAGLNSGFMGAQVTTALAEARATG-PASIH 517
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 463 ----ALAAVELV---RSKVPMDKDRHLRDIEAVS 491
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
US-09-833-745-52
; Sequence 52, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-833-745-52

Query Match 28.1%; Score 749; DB 2; Length 513;
Best Local Similarity 34.7%; Pred. No. 1.1e-64;
Matches 175; Conservative 101; Mismatches 209; Indels 20; Gaps 6;

QY 18 RHIDLDQAHAVASGGARIVLAPPARDRCRASEARLCAVIREARHVGLTTFGFLANRLI 77
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 12 RSLSLDLHRIIYEGETVGASDESMEKVKQSKRKAQVQIIVADEKLIYITGTFKFSDFI 71
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 78 SGENVRLTQANLVHLSASGVPLDWTARTAMVLAARLVLSIAQASGASEGTIARLIDLN 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 72 DPDDVENLQNLIIYSHACGVSPFPETVSRKTLVLRANALLKGFSGVRPLVIERLLALVN 131
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 138 SELAPAVPSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGTLDGAEGLRRGLQPLDLS 197
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 132 ANIHPVPIQOGLSGASGDLAPLSHLALVLLGEGEVFDYKGTTKASFALKBESIEPILTT 191
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 198 HRDALALVNGTSAMTGIALVNAHACHRLGNWAVALTALLAECLRGTEAWAAALS-DLRP 256
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192 AKEGALINGTQAMTAMGVIAVLEAEKLAFOSEIIASLTWEGRLGIIDAFDEQIHFAIRG 251
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 257 HPGQKDAARLARVRDGSARVRVHVAERLADGDTGTEPEAQDAYSRLRCAPQVIGAGF 316
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 252 YVEQVDVARRMESYLQDSQLTTSRQGLRV-----QDAYSLRCIPQVHGATW 299
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 317 DTLAWHDRVLTIELNAVTONPVPDPGSPALHGGNFMGCHVALTSDALATATVTVLAGLA 376
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 QTLRVYVKEKLEIEMNAATONPLIFDNGQV--YSGGNFHGQOQIALAMDFLGIAAEALANIS 357
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 377 ERQIARLTDERLNRGLPPLHRRGAGLNSGFMGAQVTTALAEARATG-PASIHISITN 435
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 ERRIELVNPQLNRDLPPFLSAAP-GVQSGVWILQYCAASLVSENKTLAHPASVDSIPSS 416
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 436 AANQDVSVLGTITAAKLCREKIDRWAEIILAILCALQAABEUCGSLDGVSPAGKKLVQA 495
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 417 ANQEDHVSNGTISGRHAYQIIONVRNVLAIELICAMQAVDIR---GREKMASFTKKILEK 473
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 496 LREQFPPLPPLHRRGAGLNSGFMGAQVTTALAEARATG-PASIHISITN 520
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 474 GREHVEYIDQDRMFAXKDIERAAKWL 498
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
US-09-833-745-58
; Sequence 58, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
```

```

; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 58
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Thermoplasma acidophilum
; US-09-833-745-58

Query Match          28.1%; Score 749; DB 2; Length 513;
Best Local Similarity 37.4%; Pred No. 1.le-64;
Matches 199; Conservative 91; Mismatches 194; Indels 48; Gaps 15

QY      12   PAVELD-RHIDLQAHAVASGARIVLAPPARDRCRASEARLGA---VIREARHYVGTT 67
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       5   PMIEIDGRSLRVEDVYAVEYDRVSISD--DTLKAVEEKHEAFLKILNSKTYGVNT 61

QY      68   GFCEPLANRLLSGENVRTLOANLVHLHLAGVGPVLDDWTARTAMWLARLVSIAGCSGASEG 127
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       62   GFGSLLNVNHERDQEIELQRNLIRSHSGGVGDYLENRVVRIMAVRLNSLAAGYSAVSAD 121

QY      128  TIARLIDLLNSELPAPVPSRGTVCGASGLTPLAHMVLCLQGQDFLDRDGTRLDCGAELR 187
         ::::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       122  LNNVMVEMLRNDVIPVPKYGSVCASGDLAPLANIHLAMMGEGRAFDEGRLLMSARALE 181

QY      188  RGRLOPDLDSHRDALALVNQTSAMTG---IALVNAHACRHIGNNAVALTALLAECLRGT 244
         ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       182  KAGLKPYQPKKEGVALINGTSPMSGILSIAMDAHDI--LEN-AIRSALLSFALGGTS 238

QY      245  EAWAA--ALSDLRPHPGOKDAAAARRVVDGSGARVVHVIAERRLDAGDIGTEPEAGODAY 303
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       239  KAFTPWILGALRPHLGVAIGNRFREYITGS-----DIVASKRADSVKV-----QDAY 286

QY      304  SLRCAPQVLGAGFDTLAWHDRLVTIELNAVTDNF-VFPDPDGVPALHGNGFMGOHVALTS 362
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       287  TLRCIPQVGSVADVVDYVENVLVSINSATDNPLVF--NGEEVVSNGFNHFGEPAVLA 343

QY      363  DALATAVTVLAGLAERQIARLTDERLNGLPPFLHRGPAGLNSFGMGAQVTTATALLAEMR 422
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       344  DFLAIALTDLGNMVVERIARLVDTNLSRGLPPFL-TPDSGLNSGYMI PQYTAAALCNRNK 402

QY      423  ATG-PASIHSTISWAAQDVVSLGTIAARLCREKIDRWAEITAILALCLAQAAELRCGSG 481
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       403  VLAYPSADTIPTSANQEDHSVGMATGSKLLEIIDNVRYIALEYLLQSQALEF--TD 459

QY      482  LDGVSPAGKKLVQALRRGFPLETDRPLG-----QETAAALATHL 520
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB       460  KLGNSSPTRKIYEKIREKVEKLDRDRPPSFDIETRKMWDKCKETSIALPAHL 511

```

RESULT 5

```

RESOL 3
US-09-833-745-40
; Sequence 40, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION
; TITLE OF INVENTION: BIOACTIVE HISTIDINE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770

```

RESULT 6
US-09-833-745-46
; Sequence 46, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Deinococcus radiodurans
US-09-833-745-46

```

; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 40
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-833-745-40

Query Match      27.3%; Score 728.5; DB 2; Length 508;
Best Local Similarity 35.5%; Pred. No. 1.le-62;
Matches 182; Conservative 94; Mismatches 208; Indels 29; Gaps 9

Qy 20 IDLDQAHAVASGGARIVL-----APPARDRCRASEARLGAIVREARHVGLTGTGFGPL 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2 VTLDGSLTTADVARVLDFEEAASSEMERVKSRAAVERIVRDEKTIYGINTGFGKF 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 73 ANRLISGENVRTLQANLVHLASGVGPLDWTTCARAVMLARLVIAQAGCAGSEGTIARL 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 62 SDVLIQKEDSAALQNLVILSHACGVGDPPECVSRAMLLLRANALLKGFSGVRAELTQL 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 133 IDLNSLAPAPSRGTVGASGDLTPIAHVMVLCIQGRGDFLDRDGTRLDGAEGLRGRQLQ 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 LAFLNKRVRHPVPIQOQSLGASGDLAPLSHLALALIQGEVF-PEGEMPAMTGLKKAGIQ 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 193 PLDSLHRDALALVNGTSAMTGIALVNAHACHLGNWAVALTALLAECLRGHTEAWAALS 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 PVLTLTSKEGLALNGTQMTAMGVVAYIEAKLAYQTERIASLTIEGLQGIIDAFDEDIH 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 253 DLRPHPOQKDAARLRARVDSGRVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPQVL 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 LARGYQEQIDVAERIRPYLSDSGLTTSQ--GELRV-----QDAYSLRCIPQVH 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 313 GAGFDTLIAHDRVLTITELNAVTONPVPDPGVSVALHGPNMGQHVALTSDALATATVTL 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 GATWQTGLGVYKKELEIEMNAATDNPLIFNDGD-KVISGGNFHGQPIAFAMDFLKAISEL 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 373 AGLAERQIARLTBERLNRGLPPFLHRGPAGLNSGFMGAQVATATALLAEMRATG-PASIH 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 ANTAERRIERLVNPQLN-DLPPFLSPHP-GLQSGAMTMOYAAASLVSENKTLAHPASVDS 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 432 ISTNANQDVVSGLTIAARLCREKIDRWAILATLALCQAARLCGSGLDGVSPAGK 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 IPSANQEDHVSNGTINARHAYQVIANTRRVIAIEALCALQAVEYR---GIEHAASYTKQ 460
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 492 LVQALREQFPPLTDFLQGEIQAALATHLLQQS 524
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 LFOEMRKVPSIIOODRVFSYDIERLTDLWLKES 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6
US-09-833-745-46
; Sequence 46, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Deinococcus radiodurans
US-09-833-745-46

Db 194 EGIALLNGTQASTAFALBGLFVAEDLFASATVCGAMSV-----AALGSRPPDP 243
Qy 257 -----HPGKDAARARVDSARVVRHVIARERLDAGDIGTPEAGODAYSRLCA 308
Db 244 RIHRVVRGHTQMDAAT-----AYRHLVSSEIGOSHNSCEGKV-QDPYSLRCQ 291
Qy 309 PQVLGAGPDTLAWDRVLTIELNAVTPVPFPPDGSVPALHGGNFMGQHVALTSDALATA 368
Db 292 PQVWGACIQQTRSAEVLVEANSVDNPLVPADGDI--ISGNGFHAEPVVAADNLTALA 349
Qy 369 VTVLAGLAERQIARTLDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATG-PA 427
Db 350 IAEIGSLSERMALLIDGALSCKLPFLVDN-GVNSGFMIAQVTTAALASENKTALHPA 408
Qy 428 SIHSISTWAAQNVVSLGTIAARLCREKIDRWAEILAILALCLAAQAEALRCGSLDGVP 487
Db 409 SVDSLTSANQEDHVMATFAARLRDMGENTRGILAVEYLAAAGQLDFRAPL----KSSP 465
Qy 488 AGKKLVQALREQFPPLTDRLPQEI-----AALATH 519
Db 466 RIEARQLREKVPYDKRYFADPIEKANALLAQLAVH 504

RESULT 9
US-09-833-745-56
; Sequence 56, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 56
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-745-56

Query Match 26.6%; Score 710.5; DB 2; Length 513;
Best Local Similarity 36.8%; Pred. No. 6.8e-61;
Matches 190; Conservative 82; Mismatches 222; Indels 23; Gaps 9;

Qy 10 PKPAVELDRHIDLDOAHAVASGGARIVLAPPARDRCRASEARLGAIVREARHVVYGLTTGF 69
Db 5 PEKVIELDGLTDELVN-LGKGRVYKILPTAERKRVQKSREVIDSIIEKVTWYGIITGF 63
Qy 70 GPLANRLISGENVRTLOANLVHILASGVPVLDWTTARAMVLARLVSTAOAGSASEGTI 129
Db 64 GKPAATRTVIPKQLQQLVNLVRSHSSGVGKPLSPERCRLMLARINVLAKYSGISLETL 123
Qy 130 ARLIDLNSLAPAVPSRGTVGASGDLTPLAHMWVLCQGRGDFLDRDGTLDGAEGLRRG 189
Db 124 KQVLEMFNASCPLVPYKGTGASGDLAPLSHLALGLVGEKMWPSKSGWADAKVYLEAH 183
Qy 190 RLQPLDLSHRDALVNGTSAMTGIALVNAHACHRLGNWVALTALLAECLLRGRTEAWAA 249
Db 184 GLKPVILKPEGLALINGTOMITSLGCEAVERASAIARQADIVAALTLEVLKGTTKAFDT 243
Qy 250 AL-SDLRPHQCKDAARLARVDSARVVRHVIARERLDAGDIGTPEAGODAYSRLCA 308
Db 244 DIHAPLRPHRQOI EVAFRFRSLSDSEIAESHRFCDSGRV-----QDAYTLRCC 291
Qy 309 PQVLGAGPDTLAWDRVLTIELNAVTPVPFPPDGSVPALHGGNFMGQHVALTSDALATA 368
Db 292 PQVGVVNDTIAFVKNIIITELNSATDNPMVFANGE--TVSGGNFHEGYPKALDYLAIG 349

Qy 369 VTVLAGLAERQIARTLDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRA-TGPA 427
Db 350 ITHLAAISERRIERLCPNPSLSRELPAFL-VAEGGLNSGFMIAHCTAAALVSENKALCHPS 408
Qy 428 SIHSISTWAAQNVVSLGTIAARLCREKIDRWAEILAILALCLAAQAE-LRCGSLDGVS 486
Db 409 SVDSLTSANQEDHVMGWAARVIEHVEQVLAIELLAACQGIIFLR---PLKTTY 465
Qy 487 PAGKKLVQALREQFPPLTDRLPQEI AALATHLLQQ 523
Db 466 PL-EKVVDLVRSVRPWKDRFMAPDIEAAHRLLEQ 501

RESULT 10
US-09-833-745-65
; Sequence 65, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Streptomyces griseus
US-09-833-745-65

Query Match 26.5%; Score 706; DB 2; Length 504;
Best Local Similarity 38.6%; Pred. No. 1.8e-60;
Matches 192; Conservative 69; Mismatches 212; Indels 24; Gaps 8;

Qy 27 AVASGGARIVLAPPARDRCRASEARLGAIVREARHVVYGLTTGFGLANRLISGENVRTLQ 86
Db 15 AVARHGARVELSAAVAALAAARLIVDALAAKPEPVGVSTGFGALASRHTGTETLRAQLQ 74
Qy 87 ANLVHILASGVPVLDWTTARAMVLARLVSTAOAGSASEGTIARLIDLNSLAPAVPS 146
Db 75 RNVIRSHAAGNGPRVEREVREALMFLRLKTVASGHTGVRPEVAQTMADVLNAGITPVVHE 134
Qy 147 RGTVGASGDLTPLAHMWVLCQGRGDFLDRDGTLDGAEGLRRGLQPLDLSHRDALAVN 206
Db 135 YGSLGSGDLAPLSHCLATLMGEAGEGPDGTVRPAAGELLAHAAGIAPVELREKEGLALLN 194
Qy 207 GTSAMTGIALVNAHACHRLGNWVALTALLAECLLRGRTEAWAAALSDLRPHPGKDAAR 266
Db 195 GTDGMGLMWALADLANLYTSDITAALESLEALGTDKVLAPELHAIIRPHPGQVSADN 254
Qy 267 LRARVDSARVVRHVIARERLDAGDIGTPEAGODAYSRLCAPOVLGAGFTTFLAWHDRV 326
Db 255 MSRVLAGSGLTGH- ----QDDAPRV-----QDAYSVRCAPQVNGAGRTLDHAALVA 302
Qy 327 TIELNAVTDNPFPPDGSVPALHGGNFMGQHVALTSDALATVAVTVLAGLAERQIARTLDE 386
Db 303 GRELASSVDNPNVLPDGRVES--NGNFHGPVAVVLDLFLATVAADLGSICERRTRDLRDK 360
Qy 387 RLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEM-RATGCSHTSITNAANDVVSIG 445
Db 361 NRSHGLPPFL-ADGAGVDSGLMIAQYTOAAVLSEKMLAVPASADSIPTSSAWQEDHSMG 419
Qy 446 TIAARLCREKIDRWAEILAILALCLAAQAEALRCGSLDGVS-PAGKLVQALR----BOFP 501
Db 420 WSAARKLRTAVDNLARIVAVELYAATRAIELRAAEGLT-PAPASEAVVAALRAAGAEPPG 478

Qy 502 PLETRPLGQEIHAALAT 518
Db 479 P---DREPLAPLAAADT 492

RESULT 11
US-09-833-745-41
; Sequence 41, Application US/09833745
; Patent No. 6939541
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATHARAJAN
; APPLICANT: MACALISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Streptomyces griseus
US-09-833-745-41

Query Match 26.5%; Score 706; DB 2; Length 516;
Best Local Similarity 38.6%; Pred. No. 1.9e-60;
Matches 192; Conservative 69; Mismatches 212; Indels 24; Gaps 8;

Qy 27 AVASGGARIVLAPARDRCASEARLGVAREARHVYGLTTGFGPLANRLISGNVRLT 86
Db 20 AVAHGARVELUSAAVEALAAARLIVDALAKPEFVYGVSTGFGALASRHIGTELRAQLQ 79

Qy 87 ANLVHHLASGVPVLDVTTARAVMLARLVSAQAAGSASEGTTARLDLNLSELAPAVPS 146
Db 80 RNVIRSHAAAGVPRVEREVVRLMFLRLKTVASGHTGVRPEVAQTMADVLNAGITPVVHE 139

Qy 147 RGTVGASGDLTPLAHMWLCTQGRGDFLDRTDGLDGAELGRRLGRLQPLDLSHRDALVN 206
Db 140 YGSLGSGDLAPLSHLCALTLMGEAGEBPGDGTVRPAGELLAHAAGIAPVELREKEGLLN 199

Qy 207 GTSAMTGIALVNAHACHRLGNWALVALLAECLRGRTAEWAALSLDLRPHPGOKDAAR 266
Db 200 GTDGLMGLWNLADLRLNLYTSADITAALESALGLTDKVLAPELHAIIRPHPGVGSADN 259

Qy 267 LRARVDSARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLAWHDRV 326
Db 260 MSRVLAGSGLTGH- ---QDDAPRV- ---QDAYSVRCAPQVNGAGRTLDHAALVA 307

Qy 327 TIELNAVTDNPPFPPDGSVPALHGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDE 386
Db 308 GRELASSVDNPPVLPDGRVES- -NGNFHGAPVAVYLDLAIVAADLGSICERRTRDLLDK 365

Qy 387 RLNRGLPPELHRCGAGLNSGFWGAQVATATALLAE- -RATGPASTHSITNAANOVDVSLG 445
Db 366 NRSGLPPEFL-ADGAGVDSGLMIAQYTOAALLVSEMKRLAVPASADSIPSSAMQEDHVS 424

Qy 446 TIAARLCREKIDRWAEITATIALALCIAQAELRCGSGLDGVSAGKLVQALR- ---EQFP 501
Db 425 WSAARKLRTAVDNLARIVAVELYAATRAIELRAAEGLT-PAPASEAVVAALRAAGSGPG 483

Qy 502 PLETRPLGQEIHAALAT 518
Db 484 P---DREPLAPLAAADT 497

RESULT 12
US-09-949-016-10001
; Sequence 10001, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10001
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10001

Query Match 26.1%; Score 695.5; DB 2; Length 678;
Best Local Similarity 35.7%; Pred. No. 3.2e-59;
Matches 193; Conservative 81; Mismatches 227; Indels 39; Gaps 11;

Qy 6 AMSP-----PKPAVELD-RHIDLDQAHAVASGGARIVLAPPARDRCR 46
Db 109 AMSPDFIPSPQEGVLYSKYREPEKYEIDGDRLTEDLVNLGKRYKIKLTPTAEKRVQ 168

Qy 47 ASBARLGVAREARHVYGLTTGFGPLANRLISGNVRLTQANLVHHLASGVPVLDVTTA 106
Db 169 KSREVIDSIIEKVTVVYGITTFGKFAITVPIKQLQELQVNLVRSHSSGKPLSPERC 228

Qy 107 RAMVLARLVSAQAAGSASEGTTARLDLNLSELAPAVPSRTTVGASGDTPLAHMWLCL 166
Db 229 RMLLAURINVLAKYSGISLETLLQVIEMFNASCLPYVPEKTVGASGDLAPUSHLALGL 288

Qy 167 QCRGDFLDRTDGLDGAELGRRLGRLQPLDLSHRDALALVNGTSAMTGIALVNAHACHRLG 226
Db 289 VGEKMWSPKSGWADAKYVLEAHGLKPVILKPEGLALINGTQMITSLGCEAVERASAIA 348

Qy 227 NWAVALTALLAECLRGRTAEWAALSLDLRPHPGOKDAARLRAKRVDSARVVRHVIAERR 286
Db 349 ROADIVAALTLEVLKGTTKAFDTDIHALRPHRGQIEVAFRFRSLDSDHHPSEIAESHRF 408

Qy 287 LDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLAWHDRVLTIELNAVTDNP-VFPPDGSV 345
Db 409 CD-----RVQDAYTLRCCPQVHGVNDTIAFVKNIITTELSATDPMVFNARGE- 458

Qy 346 PALHGNFMGQHVALTSDALATAVTVLAGLAERQIARLTDERLNRGLPPELHRCGAGLNS 405
Db 459 -TISGNGFHEGYPAKALDYLAIGIHELAAISERRIELCNPSLSE-LPAFL-VAEGGLNS 515

Qy 406 GFNGAQTATALLAEWRA-TGPASIHSTISNAANOVDVSLGTTAARLCREKIDRWAEILA 464
Db 516 GFMIATHTAAALVSEKALCHPSVSDLSLSTSAATEDHVSIMGWAAARKALRVIEHVEQVLA 575

Qy 465 ILALCLAAQAE-LRCGSGLDGVSAGKLVQALREQPPELTPRPLGQEIHAALTHLLQ 523
Db 576 IELLAACQIEFLR- ---PLKTTTFL- -EKVYDLVRSVVRPWIKORFMAPPDIEAHRLLLEQ 631

RESULT 13
US-09-949-016-6174
; Sequence 6174, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:05:35 ; Search time 165 Seconds
(without alignments)
1331.989 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPKPAVELDRHI.....RPLQETALNALTHLLQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	100.0	526	4	US-10-621-826-3
2	2665	99.9	526	6	US-11-069-633-8
3	2444	91.6	482	4	US-10-369-493-7816
4	1335	50.0	540	4	US-10-464-609-4
5	837	31.4	529	4	US-10-369-493-8812
6	774.5	29.0	513	3	US-09-833-745-49
7	755	28.3	501	4	US-10-369-493-4756
8	755	28.3	501	4	US-10-369-493-7515
9	755	28.3	507	4	US-10-282-122A-49837
10	753.5	28.2	501	4	US-10-369-493-19211
11	749.5	28.1	507	4	US-10-282-122A-50026
12	749	28.1	513	3	US-09-833-745-52
13	749	28.1	513	3	US-09-833-745-58
14	748	28.0	522	4	US-10-369-493-9120
15	738.5	27.7	499	4	US-10-159-257A-189
16	734.5	27.5	502	4	US-10-369-493-7044
17	734.5	27.5	511	4	US-10-282-122A-49086
18	729.5	27.3	487	4	US-10-369-493-4288
19	729	27.3	460	4	US-10-369-493-10228
20	728.5	27.3	508	3	US-09-833-745-40
21	728.5	27.3	508	4	US-10-369-493-23365
22	724	27.1	513	3	US-09-833-745-46
23	721.5	27.0	513	3	US-09-833-745-44
24	717.5	26.9	513	3	US-09-833-745-50
25	715	26.8	494	4	US-10-369-493-13631
26	715	26.8	514	4	US-10-156-761-10862
27	712	26.7	511	4	US-10-369-493-17274

28	711.5	26.7	524	4	US-10-369-493-23452	Sequence 23452, A
29	710.5	26.6	513	3	US-09-833-745-56	Sequence 56, Appl
30	709	26.6	506	4	US-10-369-493-4743	Sequence 4743, Ap
31	709	26.6	506	4	US-10-369-493-7503	Sequence 7503, Ap
32	708.5	26.6	511	4	US-10-282-122A-77245	Sequence 77245, A
33	706	26.5	496	4	US-10-369-493-18118	Sequence 18118, A
34	706	26.5	504	3	US-09-833-745-65	Sequence 65, Appl
35	706	26.5	516	3	US-09-833-745-41	Sequence 41, Appl
36	703.5	26.4	485	4	US-10-369-493-10919	Sequence 10919, A
37	699.5	26.2	484	4	US-10-369-493-8641	Sequence 8641, Ap
38	695	26.0	510	4	US-10-369-493-11500	Sequence 11500, A
39	694	26.0	497	4	US-10-369-493-7583	Sequence 7583, Ap
40	690	25.9	513	3	US-09-833-745-45	Sequence 45, Appl
41	688.5	25.8	511	4	US-10-282-122A-44428	Sequence 44428, A
42	688	25.8	507	4	US-10-369-493-15414	Sequence 15414, A
43	688	25.8	507	4	US-10-369-493-15782	Sequence 15782, A
44	688	25.8	507	4	US-10-369-493-16165	Sequence 16165, A
45	687	25.7	482	4	US-10-369-493-4824	Sequence 4824, Ap

ALIGNMENTS

RESULT 1
US-10-621-826-3
; Sequence 3, Application US/10621826
; Publication No. US20040059103A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company, Inc.
; APPLICANT: Huang, Lixuan
; APPLICANT: Xue, Zhixiong
; TITLE OF INVENTION: DNA and Amino Acid Sequences of a Tyrosine Ammonia Lyase Enzyme
; FILE REFERENCE: from the Bacterium Rhodobacter sphaeroides
; CURRENT APPLICATION NUMBER: US/10/621.826
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 526
; TYPE: PRT
; ORGANISM: R. sphaeroides
US-10-621-826-3

Query Match	100.0%;	Score	2668;	DB	4;	Length	526;
Best Local Similarity	100.0%;	Pred. No.	5.5e-214;				
Matches	526;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	VKPMAMSPKPAVELDRHIDLDQAHAVASGARIVLAPPRADRCRAEALGVAREAR	60				
Db	1	VKPMAMSPKPAVELDRHIDLDQAHAVASGARIVLAPPRADRCRAEALGVAREAR	60				
Qy	61	HVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDWTARTAMVLARLVSIAGQ	120				
Db	61	HVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDWTARTAMVLARLVSIAGQ	120				
Qy	121	ASGASEGTIARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGFLDRDGTFL	180				
Db	121	ASGASEGTIARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGFLDRDGTFL	180				
Qy	181	DGAEGLRGRLOPLDLSHRDALALVNGTSAMTGTALVNAHACRLGNWAVALTALLAECL	240				
Db	181	DGAEGLRGRLOPLDLSHRDALALVNGTSAMTGTALVNAHACRLGNWAVALTALLAECL	240				
Qy	241	RGRTEAWAALSDLRPHPGQKDAARLRARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ	300				
Db	241	RGRTEAWAALSDLRPHPGQKDAARLRARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ	300				
Qy	301	DAYSILRCAPQVLGAGFTTLAWHDRVLTIELNAVTDNVPFPPDGSVPALHCGNFMGQHV	360				
Db	301	DAYSILRCAPQVLGAGFTTLAWHDRVLTIELNAVTDNVPFPPDGSVPALHCGNFMGQHV	360				
Qy	361	TSDALATATVVIAGLAERQIARLTDRINRGLPFLHRGPAGLNSGFMGAQVTTALLAE	420				

Db 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVTTALLAE 420
Qy 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Db 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Qy 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQQSPV 526
Db 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQQSPV 526

RESULT 2

US-11-069-633-8
; Sequence 8, Application US/11069633
; Publication No. US20050208643A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt-Dannert, Claudia
; APPLICANT: Watts, Kevin
; TITLE OF INVENTION: Flavonoids
; FILE REFERENCE: 09531-160001
; CURRENT APPLICATION NUMBER: US/11/069,633
; CURRENT FILING DATE: 2005-03-01
; PRIOR APPLICATION NUMBER: US 60/549,077
; FILING DATE: 2004-03-01
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-11-069-633-8

Query Match 99.9%; Score 2665; DB 6; Length 526;
Best Local Similarity 99.8%; Pred. No. 9.8e-214;
Matches 525; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VKPWLMSPPKPAVELDRHLDLQAHAVASGGARIVLAPPARCRASEARLGAIVREAR 60
Db 1 MKPWLMSPPKPAVELDRHLDLQAHAVASGGARIVLAPPARCRASEARLGAIVREAR 60
Qy 61 HVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLDWTTARAVLARLVSTIAQG 120
Db 61 HVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLDWTTARAVLARLVSTIAQG 120
Qy 121 ASGASEGTIARLIDLINSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTTL 180
Db 121 ASGASEGTIARLIDLINSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTTL 180
Qy 181 DGAEGLRGRLOPLDLSHRDALALVNCTSAMTGTALVNAHACRHLGNWVALTALLAECL 240
Db 181 DGAEGLRGRLOPLDLSHRDALALVNCTSAMTGTALVNAHACRHLGNWVALTALLAECL 240
Qy 241 RGRTEAWAAALSDLRPHPGKDAAARLARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ 300
Db 241 RGRTEAWAAALSDLRPHPGKDAAARLARVDGSRVVRHVIAERRLDAGDIGTEPEAGQ 300
Qy 301 DAYSLRCAPOVLGAGFTFLAHDRVLTIELNAVTDNVPFPPDGSVPALHGNFMGQHVVAL 360
Db 301 DAYSLRCAPOVLGAGFTFLAHDRVLTIELNAVTDNVPFPPDGSVPALHGNFMGQHVVAL 360
Qy 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVTTALLAE 420
Db 361 TSDALATAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVTTALLAE 420
Qy 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Db 421 MRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEIILALCLAQAAELRCGS 480
Qy 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQQSPV 526
Db 481 GLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQQSPV 526

RESULT 3

US-10-369-493-7816
; Sequence 7816, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 7816
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-7816

Query Match 91.6%; Score 2444; DB 4; Length 482;

Best Local Similarity 100.0%; Pred. No. 2.5e-195; Mismatches 0; Indels 0; Gaps 0;
Matches 482; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 43 DRCRASEARLGAIVREARHVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLD 102
Db 1 DRCRASEARLGAIVREARHVYGLTTGFGPLANRLISGENVRTLOANLVHHLASGVGVLD 60
Qy 103 WTTARAVLARLVSTIAQAGSASEGTIARLIDLINSELAPAVPSRGTVGASGDLTPLAHM 162
Db 61 WTTARAVLARLVSTIAQAGSASEGTIARLIDLINSELAPAVPSRGTVGASGDLTPLAHM 120
Qy 163 VLCLQGRGDFLDRDGTTLGAGLRGRLOPLDLSHRDALALVNCTSAMTGTALVNAHAC 222
Db 121 VLCLQGRGDFLDRDGTTLGAGLRGRLOPLDLSHRDALALVNCTSAMTGTALVNAHAC 180
Qy 223 RHLGNWVALTALLAECLRGRTEAWAAALSDLRPHPGKDAAARLARVDGSRVVRHV 282
Db 181 RHLGNWVALTALLAECLRGRTEAWAAALSDLRPHPGKDAAARLARVDGSRVVRHV 240
Qy 283 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFTFLAHMHDVLTIELNAVTDNVPFPPD 342
Db 241 AERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFTFLAHMHDVLTIELNAVTDNVPFPPD 300
Qy 343 GSVPALHGNFMGQHVVALTSDALATATVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAG 402
Db 301 GSVPALHGNFMGQHVVALTSDALATATVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAG 360
Qy 403 LNSGFMGAQVTTALLAEEMRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEI 462
Db 361 LNSGFMGAQVTTALLAEEMRATGPASIIHSISTNAANQDVSVLGTIAARLCREKIDRWAEI 420
Qy 463 LAILALCLAQAAELRCGSLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQ 522
Db 421 LAILALCLAQAAELRCGSLDGVSPAGKKLVQALREQPPPLETDRPLGQEIAAALATHLLQ 480
Qy 523 QS 524
Db 481 QS 482

RESULT 4

US-10-464-609-4
; Sequence 4, Application US/10464609
; Publication No. US20040029230A1
; GENERAL INFORMATION:
; APPLICANT: KYNDT, John, Jozef Armand
; APPLICANT: VAN BEEUMEN, Jozef

; TITLE OF INVENTION: No. US20040029230A1el Methods For Synthesis of
; FILE REFERENCE: 50304/008001
; CURRENT APPLICATION NUMBER: US/10/464,609
; PRIOR FILING DATE: 2003-06-18
; PRIOR APPLICATION NUMBER: US 60/389,593
; PRIOR FILING DATE: 2002-06-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-10-464-609-4

Query Match 50.0%; Score 1335; DB 4; Length 540;
Best Local Similarity 56.5%; Pred. No. 1.3e-102;
Matches 292; Conservative 59; Mismatches 162; Indels 4; Gaps 4;

QY 11 KPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCSEARLGAIVREARHYVGLTTGFG 70
DB 19 KDCALDQALTLVQCEATATHRSRISVTPALRERCARAHARLEHAIAQRHYIGITGFG 78
QY 71 PLANRLISENVRTQANLVHHLASGVPVLDWTTTARAWLARLVSIAQAGSASEGTIA 130
DB 79 PLANRLICADQAGAEQQNLIYHLATGVGPKLSWAERARALRLNSIIQAGSASEPTID 138
QY 131 RLIDILNSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRGR 190
DB 139 RIVAVINAGFAPEVPAQTVGASGDLTPLAHMVLALQGRGMDPSGRVQEGAVMDRLC 198
QY 191 LOPDLSHRDALALVNGTSAMTGI-ALVNAHACHRLGNWAVALTALLAECLRGRTAEAWAA 249
DB 199 GGPLTLAARDGLALVNGTSAMTATAALTCVEAARAI-DAALRHSAVLMEVLSGHAEAWHP 257
QY 250 ALSDLRHPGQKDAARLARVDSGARVVRHVIARERLDAGDIGTEPEAGODAYSRLCAP 309
DB 258 APAELRHPGQURATERLAQALDAGRVCRITLTAARRLTAAADLRPDHPAQDAYSRLVVP 317
QY 310 QVLGAGFTFLAHDVRLTIELNAVTDNVPFPPDGSPALHGNFNGHVALTSDALATAV 369
DB 318 QLVGAVMTLWDHEDVVTCELSVTDNPIFEGCAVPALHGNFNGHVALASDALNAAL 377
QY 370 TVLAGLAERQIARLTDERLNRLGLPFLHRGPAAGNSFGMAQVTTALLAENRATG-PAS 428
DB 378 VTLAGLVERQIARLTDEKLNKGLPAFLHGCQAGLQSGFMGQVTTALLAENRANATPV 437
QY 429 IHSISTNAANDVVSLSGTIAARLCREKIDRWAEILAILALCIAQAALRCG-SGLDGVP 487
DB 438 VQSLSTNGANDVVSMTGTIARRARAQLPLPSIQIQAIALALALQAQMDLLDPEGQAGWSL 497
QY 488 AGKKLVQALREQFPPLTDRPLGQEIATAALATHLLQSS 524
DB 498 TARDLRDRIRAVSPGLRADRPLAGDIEVAQGLRHPS 534

RESULT 5
US-10-369-493-8812
; Sequence 8812, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8812
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Ralstonia metallidurans
US-10-369-493-8812

Query Match 31.4%; Score 837; DB 4; Length 529;
Best Local Similarity 40.9%; Pred. No. 5.2e-61;
Matches 210; Conservative 77; Mismatches 209; Indels 18; Gaps 8;

QY 19 HIDLDQAHAVASGGARIVLAPPARDRCSEARLGAIVREARHYVGLTTGFGPLANRLIS 78
DB 9 HLTPTVAIARGQPAIIVPEVLGKVADARAFQVAAANVPYIGVSTGFELVHNVWD 68
QY 79 GENVRTTQANLVHHLASGVPVLDWTTTARAWLARLVSIAQAGSASEGTIARLIDLNS 138
DB 69 IEHGRALQENLRSHCHACGVPLFSRDEVRAVMVASANALARGYSAPRPAVIEQLLKYLEA 128
QY 139 ELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEGLRGRLOPLDLSH 198
DB 129 GITPAVPQVGSIGASGDLAPLSHVAITLIGEGKVLTEGGTAPTAEVLRERGITFLALAY 188
QY 199 RDALALVNGTSAMTGIALVNAHACHRLGNWAVALTALLAECLRGRTAEAWAALS 257
DB 189 KEGALINGTSAMTGVSCULLETLRAQVQQAIEITALLALEGLSASADAFMAHGHDIAPKH 248
QY 258 PQQKDAARLARVDSGARVVRH--VIAERLDAGD---VIGTEPEAG---QDAYSRLCAP 309
DB 249 PQQIRSAANMRALLADSARLSHGCHGSELSAEMKTRAGEAKNTGT---GVPIQKAYTLRCIP 304
QY 310 QVLGAGFTFLAHDVRLTIELNAVTDNVPFPPDGSPALHGNFNGHVALTSDALATAV 369
DB 305 QVLGAVRDTLDHCAVTVVERELNSNDNPLFFEDGEL--FHGNGFNGHQQVAFAMDFLAIAA 362
QY 370 TVLAGLAERQIARLTDERLNRLGLPFLHRGPAAGNSFGMAQVTTALLAENRA-TGRPAS 428
DB 363 TOLGVVSERRNLRLSPLHNNLPAFLAAANEGLSGCFAGAQYPATALIAENRTTICSPAS 422
QY 429 IHSISTNAANDVVSLSGTIAARLCREKIDRWAEILAILALCIAQAALRCGSGLDGVS 488
DB 423 IQSVPSNGNDQVDSVSMGLIARNARRILDNNQYILALELLASCOAEL--AGAVEQLAPA 480
QY 489 GKLVQALREQFPPLTDRPLGQEIATAALATHLLQ 522
DB 481 GRAVFAFVRERVPFLSIDRYVTDDIEAMALLRQ 514

RESULT 6
US-09-833-745-49
; Sequence 49, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-833-745-49
Query Match 29.0%; Score 774.5; DB 3; Length 513;
Best Local Similarity 36.4%; Pred. No. 8.4e-56;

Db 185 LOAKEGLALLNGTQASTALALYNMFAIEDLYRTALVSGALSVDAAMGSKPFDARIHEL 244
Qy 256 PHPGKDAARLARVDSARVVRHVIARERLDAGDITGEPEAGODAYSILRCAPQVLGAG 315
Db 245 GHQOQIDAAAAYRSLLGSAINVSHADCDK-----VQDPYSLRCQPQVMGAC 291
Qy 316 FDTLAWHDRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGOHVALTSDALATATVVLG 374
Db 292 LDQWRHAANVLLLEANAVIDNPLIFPDTEGEV--LSGGNFHAEPPVAFADNLAALAAETGA 349
Qy 375 LAERQIARLTDERLNRLGPPFLHRRGPAAGLNGSGFMGAQVTTATALLAEMRATG-PASIHIS 433
Db 350 LAERIALLDATLS-GLPFFLVR-DGVNSGFMIAHTVTAALASENKTLAHPASVDSLP 407
Qy 434 TNAANQDVSLGTTAARLCREKIDRWABEILAILALCLAAELRCGSLDGVSPAGKLV 493
Db 408 TSANQEDHVSMTFAARKLGDIAENTANILSIELLAAAQGVDLRAP---HKTSPSLQAM 464
Qy 494 QALREQFPFLETDRPLGQETAAALATHLLQQSPV 526
Db 465 DTVRKOVAYELDHYPADIAAV-TRLVQNGTI 496

RESULT 9

US-10-282-122A-49837
; Sequence 49837, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykand, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49837
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49837

Query Match 28.3%; Score 755; DB 4; Length 507;
Best Local Similarity 38.8%; Pred. No. 3.5e-54;
Matches 199; Conservative 73; Mismatches 211; Indels 30; Gaps 11;
Qy 19 HIDLDQAHAVASGGARIVLAP---PARDRCRASEARLCAVIREARHVYGLTTGFGFLANR 75
Db 8 HLTLPQLRQIAREHVALQDPASHAIDACAQAVADIAA---KGEPAVGINTGFGFLAST 64
Qy 76 LISGNVRLTQANLVHHLASGVPLDWTTRARVLAARLVIAQAGSASGTTIARLIDL 135
Db 65 HIPDQLELLQNLVLSHAVGVGPMSPVVRLLIALKLSLGRHSGIRREVMDALITL 124
Qy 136 LNSELAPAVSRGVGASGDLTPLAHVVLCLQGRGDFLDRDGLDGAEGLRGRGLQPLD 195
Db 125 YNADVLPIVPKGVSGASGDLAPLAHMSAALLGVGEVFAK--GERMPATEGLALVGLKEPLT 183
Qy 196 LSHRDALALVNGTSAMTGIALLVNAHACHRLGNWAVALTALLAECLIRGRTEAWAALSDLR 255
Db 184 LOAKEGLALLNGTQASTALALYNMFAIEDLYRTALVSGALSVDAAMGSKPFDARIHEL 243
Qy 256 PHPGKDAARLARVDSARVVRHVIARERLDAGDITGEPEAGODAYSILRCAPQVLGAG 315
Db 244 GHQOQIDAAAAYRSLLGSAINVSHADCDK-----VQDPYSLRCQPQVMGAC 290
Qy 316 FDTLAWHDRVLTIELNAVTDNP-VFPDPGSPALHGGNFMGOHVALTSDALATATVVLG 374
Db 291 LDQWRHAANVLLLEANAVIDNPLIFPDTEGEV--LSGGNFHAEPPVAFADNLAALAAETGA 348
Qy 375 LAERQIARLTDERLNRLGPPFLHRRGPAAGLNGSGFMGAQVTTATALLAEMRATG-PASIHIS 433
Db 349 LAERIALLDATLS-GLPFFLVR-DGVNSGFMIAHTVTAALASENKTLAHPASVDSLP 406
Qy 434 TNAANQDVSLGTTAARLCREKIDRWABEILAILALCLAAELRCGSLDGVSPAGKLV 493
Db 407 TSANQEDHVSMTFAARKLGDIAENTANILSIELLAAAQGVDLRAP---HKTSPSLQAM 463
Qy 494 QALREQFPFLETDRPLGQETAAALATHLLQQSPV 526
Db 464 DTVRKOVAYELDHYPADIAAV-TRLVQNGTI 495

RESULT 10
US-10-369-493-19211
; Sequence 19211, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19211
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-10-369-493-19211

Query Match 28.2%; Score 753.5; DB 4; Length 501;
Best Local Similarity 39.0%; Pred. No. 4.6e-54;
Matches 196; Conservative 81; Mismatches 197; Indels 29; Gaps 9;
Qy 20 IDLDQAHAVASGGARIVLAPPARDRCRASEARLCAVIREARHVYGLTTGFGFLANRLISG 79
Db 8 LKUEILQVARNETVELSPDATRVASRALVDORVAGDTPAYGINTGFTLAEVRIDK 67

QY 80 ENVRTQANLVHHLASGVPLDWTWTARAMVRLVSLVIAQASGASGEGTARLIDLINSE 139
Db 68 KDLRLQNLILSHACGVTPLPEPARALLLRNCVNLAKYSGIRMETTIALDALMLNRD 127
QY 140 LAPAVSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGTLDGAEGLRRGLQPLDLSHR 199
Db 128 VVPVPERGSGASGDLAPLAHLALVFIGEAF-YQQRMPAKQALERAGLPVVLEAK 186
QY 200 DALALVNGTSAMTGIALVNAHACHRLGNVAVALTALLAECLRGRTAWAAALSDLRPHPG 259
Db 187 EGLALVNGTQAMCAVGTILLQRAESLADIADVAGMTLEGLLSHKPFPIEIHVRAHPG 246
QY 260 QKDAARLARVRDGSARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAGFDTL 319
Db 247 QKDVAAHLRLVDSSELVESHVNSCK-----VQDPYSLRCMPQVHGAAREGI 293
QY 320 AWHDRVLTTIELNAVTPVPPGSPVPLHGGFMGQHVALTSDALATATVTVLAGLAERQ 379
Db 294 AFRRILEVEVNSATONPLVFAD-TERIVSGGNFHGQPTISLAMDVVAMALTQLSSISRR 352
QY 380 IASLTDERLNRGLPPFLHRCGPAGLNGFMGAQVTTATALLAEMRA-TGPASIHISSTNAAN 438
Db 353 VEGVNPSSL-NLPAFLAKY-SGLNSGFMIAQVTSALVAESRVLHPASVDSIPSAGR 410
QY 439 QDVVSLGTIAARLCREKIDRWAEILAILALCLAAAEALRC-----GSLDGVSPAGKLVQ 494
Db 411 EDHVMGMTAAKGRQVSDPARSCLAIEILVAAQALDFRLPLKPGK-----ALAAVELV- 465
QY 495 ALREQPPLETDRPLQOEIAALA 517
Db 466 --RSKVPHMDKREHLRDIEAVS 486

RESULT 11

US-10-282-122A-50026
; Sequence 50026, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/257,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50026
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50026

Query Match 28.1%; Score 749.5; DB 4; Length 507;
Best Local Similarity 38.4%; Pred. No. 1e-53;
Matches 202; Conservative 74; Mismatches 213; Indels 37; Gaps 11;

QY 4 MLAMSPKPAVELDRHIDLOAHAVASGGARIVLAP---PARDRCRASEARLCAVIREAR 60
Db 1 MLTLTPGR-----LTLPQURRIARENVQIALDPSFAAIDRGAAQVADIAA---KGE 49
QY 61 HVYGLTGTGFLANRLISGENVRTTQANLVHHLASGVPVLDWTWTARAMVRLVSLVIAQ 120
Db 50 PAYGINTGFGRLASTHIPHQLELLQKNLVLSHAVGVGPEMPARPVVRLLMALKLSLGRG 109
QY 121 ASGASEGTTARLIDLINSELAPVPSRGTVGASGDLTPLAHMVLCIQGRGDFLDRDGT 180
Db 110 HSGIRRVNMDALVTLPFNADVLPLIPVKGSGASGDLAPLAHMSAVLLGIGDVFEIR-GERA 168
QY 181 DGAEGRLRRGLRLOPLDLSHRDALALVNGTSAMTGIALVNAHACHRLGNVAVALTALLAECL 240
Db 169 SAAEGLRVAGLAPLTLEAKEGLALLNGTQASTALALDNLFAIEDLYRTALVSGALSVDAA 228
QY 241 RGRTEAWAAALSDLRPHPCQKDAARLARVRDGSARVVRHVIAERRLDAGDIGTEPEAG 300
Db 229 AGSVKPPDARIHELGRHGQIDAAAYRSLLDGSAINVSH-----RDCDKVQ 275
QY 301 DAYSLRCAPOVLGAGPDTLAWHDRVLTIELNAVTPNP-VFPDPGSPVPLHGGFMGQHV 359
Db 276 DPYSLECPQVMGACLDQIRHAAGVLLIEANAVSDNPLIFPDTGVE--LSGGNFHAEPPVA 333
QY 360 LTSDALATAVTVLAGIAERQIARLTDERLNRGLPPFLHRCGPAGLNGFMGAQVTTATALLA 419
Db 334 FADNLIAIAAEIGALAEIRIALLDATLS-GUPPELVK-DGVNSGFMIAHVTAALAS 391
QY 420 EMRATG-PASIHSTISSTNAANQVSVSLGTIAARLCREKIDRWAEILAILALCLAAAEALRC 478
Db 392 ENKTLAHPASVDSLPTSANQEDHVSMTAFARKLTDIAENVANILAEILAAQGVDLRA 451
QY 479 GSGLDGVSPAGKLVQALREQPPPLETDRPLQOEIAALATHLLQQS 524
Db 452 P---HATSPALQHAKMTIRADVADVAHYDLHDHYFAPDIAVAVARRERA 494

RESULT 12

US-09-833-745-52
; Sequence 52, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-833-745-52

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2005, 09:04:50 ; Search time 11 Seconds
(without alignments)
228.969 Million cell updates/sec

Title: US-10-621-826-3

Perfect score: 2668

Sequence: 1 VKPMLAMSPPKPAVELDRHI.....RPLQGEIATLATHLLQSPV 526

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 26661 seqs, 4788334 residues

Total number of hits satisfying chosen parameters: 26661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/1/pubaa/US09 NEW PUB.pap.*
- 2: /cgn2_6/ptodata/1/pubaa/US06 NEW PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubaa/US07 NEW PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubaa/US08 NEW PUB.pap.*
- 5: /cgn2_6/ptodata/1/pubaa/PCT NEW PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubaa/US10 NEW PUB.pap.*
- 7: /cgn2_6/ptodata/1/pubaa/US11 NEW PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubaa/US60 NEW PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	461	17.3	716	7	US-11-105-259-2
2	104.5	3.9	470	6	US-10-979-821-14
3	103	3.9	566	6	US-10-467-657-3302
4	101.5	3.8	1124	6	US-10-858-730-12
5	100	3.7	397	6	US-10-467-657-9206
6	100	3.7	417	6	US-10-467-657-7708
7	99.5	3.7	1694	7	US-11-135-855-36
8	99.5	3.7	1709	7	US-11-135-855-35
9	98	3.7	628	7	US-11-147-047-45
10	95.5	3.6	2004	6	US-10-467-657-84
11	95.5	3.6	2004	6	US-10-467-657-6322
12	93	3.5	479	7	US-11-074-176-168
13	93	3.5	479	7	US-11-147-047-44
14	93	3.5	553	6	US-10-821-234-1636
15	93	3.5	776	6	US-10-925-970-3
16	92	3.4	1476	6	US-10-647-956A-4
17	91	3.4	611	6	US-10-467-657-4656
18	90	3.4	514	6	US-10-467-657-2664
19	90	3.4	7968	7	US-11-186-731-5
20	89.5	3.4	777	6	US-10-467-657-2474
21	89	3.3	557	6	US-10-512-109-9
22	88.5	3.3	396	7	US-11-061-869-11
23	88.5	3.3	417	6	US-10-467-657-8450
24	88.5	3.3	934	6	US-10-858-730-8
25	87.5	3.3	357	7	US-11-055-822-734

26	87.5	3.3	772	6	US-10-858-730-77	Sequence 77, Appl
27	87.5	3.3	4384	6	US-10-821-234-1120	Sequence 1120, Ap
28	87	3.3	745	6	US-10-858-730-224	Sequence 224, App
29	87	3.3	745	7	US-11-055-822-226	Sequence 226, App
30	87	3.3	745	7	US-11-055-822-714	Sequence 714, App
31	87	3.3	745	7	US-11-055-822-756	Sequence 756, App
32	87	3.3	759	6	US-10-467-657-2722	Sequence 2722, Ap
33	86.5	3.2	331	6	US-10-432-483-25	Sequence 25, Appl
34	86.5	3.2	407	7	US-11-082-389-48	Sequence 48, Appl
35	86.5	3.2	459	6	US-10-467-657-3092	Sequence 3092, Ap
36	86	3.2	436	6	US-10-467-657-7694	Sequence 7694, App
37	85.5	3.2	2280	7	US-11-022-562-211	Sequence 211, App
38	85	3.2	366	6	US-10-432-483-24	Sequence 24, Appl
39	85	3.2	474	7	US-11-055-822-432	Sequence 432, App
40	85	3.2	475	6	US-10-509-464-8	Sequence 43, Appl
41	85	3.2	653	7	US-11-137-465-55	Sequence 55, Appl
42	85	3.2	717	6	US-10-467-657-8056	Sequence 8056, Ap
43	85	3.2	3717	6	US-10-821-234-1076	Sequence 1076, Ap
44	84	3.1	1001	6	US-10-467-962B-81	Sequence 81, Appl
45	84	3.1	1389	6	US-10-467-657-334	Sequence 334, App

ALIGNMENTS

RESULT 1
US-11-105-259-2
; Sequence 2, Application US/11105259
; Publication No. US20050260724A1
; GENERAL INFORMATION:
; APPLICANT: Ben-Bassat, Arie
; APPLICANT: Sariasi, Pateme Sima
; APPLICANT: Huang, Lisa L.
; APPLICANT: Patnaik, Ranjan
; APPLICANT: Lowe, David J.
; TITLE OF INVENTION: Microbial Conversion of Glucose to Para-Hydroxystyrene
; FILE REFERENCE: CL2598 US NA
; CURRENT APPLICATION NUMBER: US/11/105.259
; CURRENT FILING DATE: 2005-04-13
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Rhodotorula glutinis
US-11-105-259-2

Query Match	17.3%	Score 461	DB 7	Length 716
Best Local Similarity	29.3%	Pred. No. 2e-28		
Matches 163	Conservative	78	Mismatches 265	Indels 50
Gaps 17				
Qy	1	VKPLAMSPPKPAVELDRH-IDLDQAHAVASGGARIVLAPPAR----	DRCRASEARLCAV	55
Db	48	VERKLA-APTDSITLEDGYSINLGDVVSARKGR-----	PVRVKDSDEIRSKDKSVEF	100
Qy	56	IRE--ARHVVGLTGTGFCPLANRLISGNVRLQANLVHHLASGVGP-----		99
Db	101	LRSQLSMVGVTTGFGGSAD--TRTEDIASLQKALLEHQLCGVLPSSFDSFRLGRGLEN	158	
Qy	100	VLDWTTARAMVLARLVSIAGCASCSEGTARLDLIDNSELAPVPSRGTVGASGDLTPL	159	
Db	159	SLPLEVVRGAMTIRVNSLTRGHSVRLVLEALTNFLNHGITTPIVPLRGTTISASGDISPL	218	
Qy	160	AHVVLCQGRGD---PLDRDGTTR--LDGAEGLRRGRLOPLDLSHRDALAVNGTSAMTGI	214	
Db	219	SYTAALISGHPDSEKHVVHVEGKEKILYARFAMALFNLEPVVLGPKGGLVNGTAVASAM	278	
Qy	215	ALVNAHACRHLGNWAVALTALLAECLFGRTEAAWAAALSDL-RPHFGQKDAARLARVDG	273	
Db	279	ATLALHDAHMLSLLSQSILTAMTVMVGHAGSFHPLHDVTRPHTQIEVAGNIRKLEG	338	
Qy	274	SARVVRHIVARRRIDAGDITGEPEAGODAYSRLCAPQVLGAGFTTLAWHDRLTIEL-NA	332	

Db 339 SRFAVHH---EEVVKVD--DEGILRQDRYPLRTSPQWLGPVLSDDLHHAHVLTIENGQS 393
Qy 333 VTDNPVPPDGSVPALHGGFMGQHVALTSDALATAVTVVLGALAEARQIARLTDERLNRGL 392
Db 394 TTDNPLDVENKT-SHEGNGFQAAAVANTWEKTRGLAQIKLNFQTQTEMLNAGMNRGL 452
Qy 393 PPFLHRGPAGLNSGFMGAQVATATALLAEM-RATGPASIHISINNAQDVVSIGTTIARL 451
Db 453 PSCLAABDPSSLVSHCKGLDAAAAYTSGLHLANPVTTHVQPAEMANQAVNSLALISARR 512
Qy 452 CREKIDRWAETLALCLAQAAELRCGSLDGVSPAGKKLQALREOPPPLETDRLPGQ 511
Db 513 TTESNDVLSLLATHLYCVLQAIIDL-AIEEFKQFGPAIVSLIDQFGSAMTGSNLRD 571
Qy 512 EIA-----ALATHLLQ 522
Db 572 ELVEKNKNTLAKRLEQ 587

RESULT 2

US-10-979-821-14
; Sequence 14, Application US/10979821
; Publication No. US20050244937A1
; GENERAL INFORMATION:
; APPLICANT: ABRAHAM, TIMOTHY W.
; APPLICANT: CAMERON, DOUGLAS C.
; APPLICANT: HICKS, PAULA M.
; APPLICANT: MCFARLAN, SARA C.
; APPLICANT: MILLIS, JIM
; APPLICANT: ROSAZZA, JACK
; APPLICANT: ZHAO, LISHAN
; APPLICANT: WEINER, DAVID P.
; TITLE OF INVENTION: POLYPEPTIDES AND BIOSYNTHETIC PATHWAYS FOR THE
; FILE REFERENCE: 023829-0390
; CURRENT APPLICATION NUMBER: US/10/979,821
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: 10/422,366
; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 60/374,831
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin ver. 3.3
; SEQ ID NO 14
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Rhodobacter sphaeroides
US-10-979-821-14

Query Match 3.9%; Score 104.5; DB 6; Length 470;
Best Local Similarity 25.2%; Pred. No. 0.29;
Matches 117; Conservative 31; Mismatches 152; Indels 165; Gaps 22;
Qy 1 VKWM-----LAMEPPKPAVEL--DRHIDLQAHAVASGGARIVLAPPARDRCRASEARLG 53
Db 98 VEPVAMRAGLAFSDGAPDPELVPDK-----ALAPAFRALLSPAFRAGADYGDARGT 149
Qy 54 AVIREARHVYGLTTGFGPLANRLISGENVRLQANLVHILASGVGVPLDWTITARAMVLAR 113
Db 150 SSUREA-----LAAYLASRGGVAD-----PAR 172
Qy 114 LVSIAGASGASEGTIARLIDLINSELAP----AVPSRGTGVASGDITPLAHMVLCLQG- 168
Db 173 LL-LARGSQMA-----LFLVARAALAPGEATAVEEPG-----YPLAWAEFRAAGA 216
Qy 169 --RGDFLDRDGTRLDGAE-----GURRGLQPLDLSHRD 200
Db 217 EVRGVPPVDGGLRIDALEAALARDPRIRAVYVTPHHQYPTVTVTWGAAR-RLQLELAERH 275
Qy 201 ALALNV-----GTSAMTGIALVNAHACRHLGNWAVAL-TALLAECLRG 242
Db 276 RLALIEDDYHREYRFEGRPVLPAAARAPEGLPIIYVGSLSKLJSPGIRUGYALAPERLLT 335

Qy 243 RTAWAAALSDLRPHFGQKDAARLARVDGS-----ARVVRHVIAERRLD----- 288
Db 336 RMAAARAAIDRQDAPLEAALAEIR---DGDIGRHAKARVYRARR-DLLAERLTAQL 391
Qy 289 AGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLAWHDRVLTIELNAVTDNVPVPPDGSVPAL 348
Db 392 AGRAADFAPAGLALWIRC-----AGVSAETWAEAA-----CGAGLALL 430
Qy 349 HGGNFMGQHVLTSDALATAVTVVLGALAEARQIARLTDERINRGLP 393
Db 431 PGTRF-----ALESAPQAFRLGYAALDEGQIARAV-EILARSFP 469
RESULT 3
US-10-467-657-3302
; Sequence 3302, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 3302
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-3302

Query Match 3.9%; Score 103; DB 6; Length 566;
Best Local Similarity 20.6%; Pred. No. 0.48;
Matches 116; Conservative 56; Mismatches 185; Indels 206; Gaps 24;
Qy 21 DLDQAHAVASG-----GARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGPLA 73
Db 82 DADGATACAVGLDGLAAMGAKVDLVPNR-----FEHGYGLTPELAEIA 125
Qy 74 -----NRLISGEN-----VRTLQANLV---HHL----- 93
Db 126 AAQGVLLITVDNGIASIAGVARAQALGLDIVTDHLPADTVPCIIIVNPNGKCGFPS 185
Qy 94 --ASGVGPVLDWTITARAMVLARLVSTIAGASGASEGTIARLIDLINSELAPAVPSRGTVG 151
Db 186 KSLAGVGVIYVYVLTALRAELRRNRYE--SDGIKEPNLGGLLDLV-----ALG 230
Qy 152 ASGDLTPLAHMVLCLQGRGDFLDRDTRLDGAEL-----RRG--RLQPLDLSHRDALALV 205
Db 231 TVADVVPVLDHNNRILVSQGLKRMRSCKMRPGIRALFEVARRDWRKAQPPDM----- 281
Qy 206 NGTSAMTGIAL-VNAHACRHLGNWAVALTALLAECLRGTEAWAAALSDLRPHPGOKDAA 264
Db 282 -----GFALGPRINAAGRLLDMSVGIACLLA---RDDSEA-----OELA 317
Qy 265 ARLRARVDGSARVVRHVIAERLIDAGDIGTEPEAGQDAYSLRCAPQVLGAGFDTLA---- 320
Db 318 ARL-----NNLNIERR-----EIEQSMRLDALNAFPETLPSGGTTLVAYRD 358
Qy 321 -WHDRVLTTELNAVTDNVPVPPDGSVPALHGGFMGQHVALTSDALATATVVLGALAEARQ 379
Db 359 DFHQGVVGVIVASRLKDRFYRPTIVFAPA-DNGEVRGSGRSIPNLHLRDALDLVSKRHPDL 417
Qy 380 IARLTDERLNRGLPPPLHRGPAGLNSGFMGAQVATATALLAEARATGPASIHISITNAANQ 439
Db 418 ILKFGGHAAAGUSILEHNIPA-----FQTAEEAVREM-----VCEDDLSQ 459

QY 440 DVVSLGTIAARLCKRIDRWAEILAILALCLAAELRC---GSLDGVSPAGKLVQAL 496
Db 460 TYITDGSUPA-----CDITLQEAQNLACHVWGQGF----- 489
QY 497 REQFPFPLETD-----RPLQGE 512
Db 490 ---APPSFTDFHVVRQQLGAE 509
RESULT 4
US-10-858-730-12
; Sequence 12, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; TITLE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1124
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-858-730-12

Query Match 3.8%; Score 101.5; DB 6; Length 1124;
Best Local Similarity 21.9%; Pred. No. 1.6; Mismatches 226; Indels 203; Gaps 33;
Matches 137; Conservative 61

QY 3 PMLAMSPKPAV-ELDRHIDLQ-----AHAVASGGARIV--LAPPARDRCRASEARLGA 54
Db 130 PVLGSSAPSTDVDELVRAD-DVGFPVVKAVAGGGGGRVVEEPAQLR-----EA-IEA 183
QY 55 VIRARHYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARL 114
Db 184 ASREAASAFGDSVTF--LEKAVVEPRHIE-----VQILDGEGDVHLHFERDCGVORRH 235
QY 115 VSIAQAGSASEGVIARLIDLNLSELAPVPSRGTGASGDLTPLAHVLCLO----- 167
Db 236 QKVI-----ELAPA-----PNLDPALRERICADAVNFAQ 265
QY 168 -----GRGDPL-DRDGTLL-----DGAGLRGRRLQPLDLSHRDALALVNGTSAMTG 213
Db 266 IGERNAGTVFLVDGDNHVFIEPNPQIYVETHTVEETVDVLVQSQRLEAAGTLLADLG 325
QY 214 IALVNAHACHRLGNWVALTALLAECLRGRTTEAWAALSRLRPHPGQKDAARLARVDG 273
Db 326 LAQEN-----ITLGAALQC-RITTEDPA---NGRPDTGQISA-----YRSPG 365
QY 274 SARVVRHVIAERRLDAG--DIGTPEAGQDAYSRLCAPQVLGAGDTPLAHWDRVLITIE-- 329
Db 366 GSGI-----RLDGGTTHAGTETSAHFDSMLVKLSGR--GRDFTTAVNARRAVAEFR 415
QY 330 -----LNATVNDVPFPPDGSPALHCGNFMQO--HVALTSDALATATVTLAGLAE 377
Db 416 IRGVATNIPFLQAVLDDPDFQA--GRVTT-----SFIEQRPHLLTARHSADRGTKLLTYLAD 470

QY 378 RQIARLUTDER-----LNRGLPPFLHR-----GPAGL-----NSGFMAQVTA 414
Db 471 VTNKPHGERPELVDPULKUTASAGBPPAGSROLLAELGEGFPARRLRESSTIG--VTD 528
QY 415 TALLAEMRATGPASISHSISTNAANQDVVSLGTIAARLCREKI----- 456
Db 529 TTF-----RDAHQSLLATVTRTKDMLAVAPVVRTLPQLLSLECWGGATYDVVALRELA 581
QY 457 -DRWAEILAILA-----LCLAQAELRCGSLDGVSPAGKLVQAE----- 495
Db 582 EDPWERLAALREAVPNLCLOM---LLRGRNTVGYTPYPTVTEVDAFVQEAATGIDIFRIF 638
QY 496 -----LREQFPFPLETDRPLQGEIQAALA 517
Db 639 DALNDVEQMRPAIEAVRQTSVAEVA 665
RESULT 5
US-10-467-657-9206
; Sequence 9206, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 9206
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-9206

Query Match 3.7%; Score 100; DB 6; Length 397;
Best Local Similarity 22.1%; Pred. No. 0.51; Mismatches 55; Indels 90; Gaps 17;
Matches 78; Conservative 55

QY 70 GPLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARL-----VSIAQAGSAS 125
Db 81 GTVNNTCAPYELVTRWEASIL-----VLGP-----TLARFGEAQVSLPGCAIGS 125
QY 126 EGTIARL--IDLNLSELAPVPSRGTGASGDL--TPLAHVLCLOGRGDFLDRDGTLLD 181
Db 126 RPVDQHLKGLEAMGARI---VIEHGYVVKAGKLGKTRVAMDVVTVGGTENLL-MAATLAE 181
QY 182 GAGLRGRRLQPLDLSHRDALALVN-----GTSMTGIALVNAHACHRLGNWVALTAL 235
Db 182 GTTVLENCALEPEVVDLAECLVFMGAKISIGISTMTLVEGVDELHGCHEH-----SV 232
QY 236 LAECLRGRTTEAWAALSRLRPHPGQKDAARLARVDSARVVRHVIAERRLDAGDIGTE 295
Db 233 VPDRIEAGTFLCAVIT-----GGRVLR--NAAPKTMEVVDLKLVEAGAV--- 276
QY 296 PEAGQD--AYSLRCAPO-----VLGAGFDT-----LAWHRVL 326
Db 277 IEAGDWDIAIDMRQRPKAVDIRTVVHGFPTDMQAFMALNAVAEGSCRVVVETTFENRFM 336
QY 327 TI-ELNAVTDNVPFPPDGSPALHCGNFMQOHWALTSDALATATVTLAGLAER 378
Db 337 HYPELNRMGAN--ITTEGNTAFVQGVRLSGAVVVKATDLRASLSLVIAGLAAR 387

RESULT 6
US-10-467-657-7708
; Sequence 7708, Application US/10467657

```
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 7708
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-7708

Query Match      3.7%; Score 100; DB 6; Length 417;
Best Local Similarity 22.1%; Pred. No. 0.55;
Matches 78; Conservative 55; Mismatches 130; Indels 90; Gaps 17;

QY 70 GPLANRLISGENVRTLQANLVHHLASGVGVLDWTTTARAMVLARL-----VSIAQGASGAS 125
DB 77 GTVNNTCAPVELVTRWEASIL-----VLGP-----TLARFGEAQSLSLPGGCAIGS 121

QY 126 EGTIARL--IDLNSELAPVPSRGTVGASGDL--TFLAHMVLCLQGRGDFLDRDGTRLD 181
DB 122 RPVDQHLKGLGEMGAET---VIEHGYVYKAGKLGKTRVAMDVVTVGGTENLL--MAATLAE 177

QY 182 GAELGRRLQPLDLSHRDALALVN-----GTSMTGIALVNAHACHLGNWAVALTAL 235
DB 178 GTTVLENCAIEPEVVDLAELCLVMGAKISIGTSTMIVEGVDELHGEH-----SV 228

QY 236 LAELGRGRTEWAAALSDLRPHPGQDMAARLARVDGSAVRVHVIAERRLDAGDIGTE 295
DB 229 VPDRIEAGTFLCAVAIT-----GRRVLR--NAAPKTMVEVLDKLVEAGAV--- 272

QY 296 PEAGQD--AVSLRCAPO-----VLGAGFDT-----LAWHDRVLL 326
DB 273 IEAGDDWIAIDMRQPKAVDRTVWHPGPTDMQAFMALNAVAEGSCRVTETIFENRFM 332

QY 327 TI-ELNAVTDNPPVPPDGSVPALHGGNFMGQHVALTSDALATVTVLAGLAER 378
DB 333 HYPELARMGAN--ITTEGNTAFVQGVVERUSGAVVKATDLRASASLVIAGLAAR 383

RESULT 7
US-11-135-855-36
; Sequence 36, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-36

Query Match      3.7%; Score 99.5; DB 7; Length 1709;
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-36

Query Match      3.7%; Score 99.5; DB 7; Length 1694;
Best Local Similarity 23.1%; Pred. No. 3.8;
Matches 123; Conservative 41; Mismatches 215; Indels 153; Gaps 24;

QY 48 SEARLGVIREARHVVYGLTTGFGPLANRL--TSGENVRTLQANLVHHLASG----- 96
DB 431 SEPLATLVLSHGHIILASTSGSDSDHSPRFGTSGPNSLRLEIRDLRETSGEYKCSATNS 490

QY 97 ---VGPVLDW--TTARAMVLARLVSTAAQG-----ASGASEGTIARLIDLIN----- 137
DB 491 LGNATSTLDFHANAAARLLISPAAEVVEGQAVTLCSCFSLSPPTDFARFSWYNGALLHEGP 550

QY 138 ---SELAPAVPS-----RGTVGASGDLTFLAHMVLCLQGRGDFLDRDGTRLD----- 181
DB 551 GSSLLLPAAASSTDAGSYHCRARDGHSASGSPSPAVLTVLYPPRQPTFT-----TRLDLDA 606

QY 182 GAELGRRLGRL-----QPLDLSHRD--ALALVNGTSAWT-----GIA 215
DB 607 GAGARRGLLRCVSDPPARLQLLHKDRVWATSLSPSGGGCSTCGGCSPRMKVTKAPNLL 666

QY 216 LVNAH-----ACRHLGNWAVALTALAECLRGRTTEWAAALSDLRPHPGQKD 262
DB 667 RVEIHNPLLEEGLYLCEASNALGNASTSAT-----FNGQATVLAAPS-----HTLQEG 716

QY 263 AARLARVVDGSA-----RVVRHVIAERRLDAGDIGTEPEAGQDAYSLRCAPOVLGAG 315
DB 717 TEANLTCNYSREAAGSPANFSWFRNGVLWAOGPLETVTLPLPVARTDAALYAC----- 768

QY 316 FDTLAWHDRVLTIELNAVTDNPV-----PPDGSVPALHGGNFMGO--HVAL---TSDALA 366
DB 769 -----RLT-EAGAQLSTVLLSVLYPPDR--PKLSALLDMGGHFWALFCTVDVSRP 817

QY 367 TAVTVLAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTTATALLAEMRATGP 426
DB 818 LALLALFH-GEHLA-----TSLGPQVPVSHGRFOAKAENSCLKLEVRELGL 862

QY 427 ASIHSISTNAANQDVSVLSGTIAARLCKEIKDRWAEILAILALCLQAABEIRC 478
DB 863 GDSGSYRCATN-----VLGSSNTSLFFQVRGAVVQVSPSPQLQEGQAVVLS 910

RESULT 8
US-11-135-855-35
; Sequence 35, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1709
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-35

Query Match      3.7%; Score 99.5; DB 7; Length 1709;
```

	Best Local Similarity	23.1%;	Pred. No.	3.9;						
	Matches	123;	Conservative	41;	Mismatches	215;	Indels	153;	Gaps	24;
QY	48	SEARLGAVIREARHYVGLTTGFGFLANRL--ISGENVRTLQANVLHLAGS-----	96							
DB	431	SEPLATLVLSHGGHILASTSGDSHPSCFSTGSPNSLRLEIRDEETDSEYKCSATNS	490							
QY	97	---VGPVLDW--TTRAMVMVARLVIAQQ-----ASCASEGTIARLTIDLN-----	137							
DB	491	LGNATSTLDFHANNAARLLISPAAEVEGQAVTLSCRSLSTPPDPARPSWYLNGALLHEGP	550							
QY	138	--SELAPAVPS-----RGTGVAGSDULTPLAHMVLCIQGRGDFLDRDGRDL---	181							
DB	551	GSSILLPAASSTDAGSYHCARDCHGSASPSSPAVLTVLYPPRPQTPT---TRLDLDA	606							
QY	182	GAELRRGRLL-----QPLDLSHRD---ALALVNGTSMT-----GIA	215							
DB	607	GAGAGRGLLCLRVDSPPARQLQLHKDRVVATSLPSCGGCGSTCGCSPRMKVTKAPNLL	666							
QY	216	LVNNH-----ACEHLGNWAVALTALLAECLRGETEAWAALSRLRHPPGOKD	262							
DB	667	RVTIHNPFLREEGYLCEASNALGNASTSAT-----FNGQATVLAITAPS----HTLQEG	716							
QY	263	AAARLARVDGSA-----RVVRHVIAERLDDAGDIGTEPGADQAYSLRCAPQVLGAG	315							
DB	717	TEAMLTCNVREAGSAPANSFWPRNGVLMAQGPLETVTLLEVARTDAALYAC-----	768							
QY	316	FDTLAWHDRVLTIELNAVTONPV-----FPDGSVPALHGFNFMGQ-HVAL---TSDALA	366							
DB	769	-----RIILT-EAGAQLSLPVLLSVLYPPDR--PKLSALLDMGQHMALFICTVDSRP	817							
QY	367	TAVTVLAGLAERQIARLTDERLNRLGPPFLHRGPAGLNSGFMGAQVATATALLAEMRATGP	426							
DB	818	LALLALPH-GEHLA-----TSLGQPVPFHGRFFQAFAEANSCLKLEVRELGL	862							
QY	427	ASTHSISTNANOQDVSLGTTAARLCREKIDRWAEIIAILALCALCAQAABURC	478							
DB	863	GDGSGSYRCEATN-----VLGSNTSLFFOVRGAWVOVSPSPBLOGOAIVLWSC	910							

```

RESULT 10
US-10-467-657-84
; Sequence 84, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 84
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-84

```

Query Match	3.6%;	Score 95.5;	DB 6;	Length 2004;
Best Local Similarity	23.3%;	Pred. No. 9.7;		
Matches 101; Conservative	45;	Mismatches 170;	Indels 117;	Gaps 21;

```

RESULT 9
US-11-147-047-45
, Sequence 45, Application US/11147047
, Publication No. US2005026068A1
, GENERAL INFORMATION:
, APPLICANT: Agarwal, Pankaj
, APPLICANT: Murdock, Paul R.
, APPLICANT: Rivzi, Safia K.
, APPLICANT: Smith, Randall F.
, APPLICANT: Xiang, Zhaoyang
, TITLE OF INVENTION: NOVEL COMPOUNDS
, FILE REFERENCE: GP50016
, CURRENT APPLICATION NUMBER: US/11/147,047
, CURRENT FILING DATE: 2005-06-07
, PRIOR APPLICATION NUMBER: US/10/221,097
, PRIOR FILING DATE: 2002-09-06
, PRIOR APPLICATION NUMBER: PCT/US01/07143
, PRIOR FILING DATE: 2001-03-05
, PRIOR APPLICATION NUMBER: 60/187,107
, PRIOR FILING DATE: 2000-03-06
, PRIOR APPLICATION NUMBER: 60/236,874
, PRIOR FILING DATE: 2000-10-03
, PRIOR APPLICATION NUMBER: 60/188,916
, PRIOR FILING DATE: 2000-03-13
, PRIOR APPLICATION NUMBER: 60/237,846
, PRIOR FILING DATE: 2000-10-03
, NUMBER OF SEQ ID NOS: 52
, SOFTWARE: FastSeq for Windows Version 3.0
, SEQ ID NO 45
, LENGTH: 628
, TYPE: PRT
, ORGANISM: Homo sapiens
US-11-147-047-45

```

QY 25 AHAVASGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGFLANRLISGENVRT 84
DB 341 SEAVVGGRLTRALPTEDKGLAQDVR-----QDVQ--GLTGGGRLTPDAGADANAAA 393
QY 85 LQANLVHHLASGVPVLD---WTTARAMVLARLVISIAQGSASEGTIA-----RLI 133
DB 394 LQGLPGSAVASGNAPARRQNQLQVRARAEGAAPGLSASENLAGTDGGRAPVAGKRPDTVL 453
QY 134 DLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFL-----DRDGTLDGA 183
DB 454 PVLNPOVAE-----SAGRVSPKKRMA---DAAADFTRLAADRREPEKAGVPLGGG 501
QY 184 EGLRRGRLOPLDLSHRDALALVNGTSAMTGI-----ALVNAHACRHL-GNW 228
DB 502 E-----YRFEHTDRRHIDALAGVPRGKGMPEPADMAGPNSDGLV-SDGERYLGRE 556
QY 229 AVALTA-LIAECL-----RGRTEMAAALSDLRPHPGQKDAARLRAR-----V 271
DB 557 AETLRAGGLSEAVPSEPRGRDYRPTQEARAPAKVMARPRDAAADGKPAQAQPARAKDTPV 616
QY 272 DGSARVVHVIAE-----RLDAG---DIGT-EPEAGQDAYSLRCAPOVLGAGFDTL 319
DB 617 AGKAAAKNAATEKPSDKVRIEAGKSRFDGKGKSAAGAAATEKPSKTKAKPETF 676
QY 320 AWHDRVLTIELNAVTNPVPPDPGSPALHGG-----NFMQGHVALTSDAL 365
DB 677 A-----KTASDNPE-EARRKARVLQGGPVVYTKERQAPQGFKALREHAESIKKRL 725
QY 366 ATAVTVLAGLAER 378
DB 726 AESI---GGLAER 735

RESULT 11
US-10-467-657-6322
; Sequence 6322, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 6322
; LENGTH: 2004
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-6322

Query Match 3.6%; Score 95.5; DB 6; Length 2004;
Best Local Similarity 23.3%; Pred. No. 9.7;
Matches 101; Conservative 45; Mismatches 170; Indels 117; Gaps 21;
QY 25 AHAVASGARIVLAPPARDRCRASEARLGAVIREARHVYGLTTGFGFLANRLISGENVRT 84
DB 341 SEAVVGGRLTRALPTEDKGLAQDVR-----QDVQ--GLTGGGRLTPDAGADANAAA 393
QY 85 LQANLVHHLASGVPVLD---WTTARAMVLARLVISIAQGSASEGTIA-----RLI 133
DB 394 LQGLPGSAVASGNAPARRQNQLQVRARAEGAAPGLSASENLAGTDGGRAPVAGKRPDTVL 453
QY 134 DLINSELAPVPSRGTVGASGDLTPLAHMVLCLQGRGDFL-----DRDGTLDGA 183
DB 454 PVLNPOVAE-----SAGRVSPKKRMA---DAAADFTRLAADRREPEKAGVPLGGG 501

QY 184 EGLRRGRLOPLDLSHRDALALVNGTSAMTGI-----ALVNAHACRHL-GNW 228
DB 502 E-----YRFEHTDRRHIDALAGVPRGKGMPEPADMAGPNSDGLV-SDGERYLGRE 556
QY 229 AVALTA-LIAECL-----RGRTEMAAALSDLRPHPGQKDAARLRAR-----V 271
DB 557 AETLRAGGLSEAVPSEPRGRDYRPTQEARAPAKVMARPRDAAADGKPAQAQPARAKDTPV 616
QY 272 DGSARVVHVIAE-----RLDAG---DIGT-EPEAGQDAYSLRCAPOVLGAGFDTL 319
DB 617 AGKAAAKNAATEKPSDKVRIEAGKSRFDGKGKSAAGAAATEKPSKTKAKPETF 676
QY 320 AWHDRVLTIELNAVTNPVPPDPGSPALHGG-----NFMQGHVALTSDAL 365
DB 677 A-----KTASDNPE-EARRKARVLQGGPVVYTKERQAPQGFKALREHAESIKKRL 725
QY 366 ATAVTVLAGLAER 378
DB 726 AESI---GGLAER 735

RESULT 12
US-11-074-176-168
; Sequence 168, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kjaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-168

Query Match 3.5%; Score 93; DB 7; Length 479;
Best Local Similarity 22.1%; Pred. No. 2.3;
Matches 62; Conservative 36; Mismatches 104; Indels 78; Gaps 11;
QY 271 VDGSARVVV---HVIARRLDAGDIGTEPEAGQDAYSLR-----CAPOVLGAGFD 318
DB 27 INNALRVKSEDESIVLEVTLELGDGLVLTIAWESTDGLRGMKVEDTGAPISVPVGEDT 86
QY 319 LAWHDRVLTIELNAVTNDNVPFPPDGSVPALHGGNFMQGHVALTSDALATATVTVLAGLAER 378
DB 87 LG---RVENVLQPIDGGPAFPKDHPEIGHKEAPKYEDLTTSREILETGKVI----- 137
QY 379 QIARLTDERLNRGLPFLHRGPAGLNSGFWAQVTTALLAENRATGPASIHSTNAAN 438
DB 138 -----DLLEPYVRGGKVGL---FGGAGVGKTTIQEL-----IHNI---AQE 173
QY 439 QDVVSGTIGTIAARLCREKIDRWAE-----ILAILALCLAAQAAELRCGSLDGVSPAGKVLV 493
DB 174 HGGISVFTGVGRTREGNDLIYEMKASGVLSKTMVFGQWNE-----PPGARMR 222
QY 494 QA-----LREQFPPLTDRPL-----GQBI AAL 516
DB 223 VALTGUTLAEYFRDVEGDVLLFIDNIPRFTQAGSEVSAL 262

RESULT 13

